Classifiers and Strategic Pricing

Rajiv M. Dewan, Marshall Freimer, Simon Business School, University of Rochester

April 2, 2019

Abstract

We look at the interplay between machine learning technology and strategic pricing. Cost of errors in classification play a significant role in building and tuning machine learning systems for use in business. These costs depend on the business situation at hand, and are typically asymmetric, i.e., the cost of a False-Positive may not be the same as the cost of a False-Negative error. This has implications for choice of re-sampling techniques, the cut-off used for classification, and indeed for the choice of the learning model itself. This gets even more complicated when we consider strategic situations involving more than one decision maker. Using the screening game, we show that strategic interactions dramatically change the costs of errors faced by decision makers, even to the point where a strategic player may benefit from errors. We next show that even when errors are costly, a consumer may use an inferior classifier, i.e, a classifier with higher rates for all error types, over another. This calls into question a vast array of literature in data science that assumes the inferior classifiers have no utility.

1 Introduction

As Machine Learning techniques find wide use in business, data scientists are grappling with issues that are particular to business settings. Chief among these is the asymmetric cost of False-Positive (Type I) and False-Negative (Type II) errors [1]. In response to this asymmetry, data scientists have developed a variety of techniques that provide asymmetric learning, i.e., the ability to reduce one kind of error at the expense of another. Resampling the data to over represent or to increase the weight of observations in the class whose errors are more costly is a common approach [2]. Ensemble learning methods such as boosting achieve a similar effect [3]. Another technique is to change the cutoff for determining the class by looking for a cutoff at which the *Receiver Operating Characteristic* curve is tangent to a straight line that reflects total costs [2]. These asymmetric machine learning techniques can allow a decision maker to tailor the predictive algorithm to favor one kind of error over another.

All of these techniques depend on the estimation of the costs of error. While this may be difficult even with only one decision maker, it gets much more complicated when more than one strategic decision maker is involved and the errors affect them differently. Perhaps, it is intuitive to see this in the case of a zero-sum game in which one strategic decision maker's gain is the other's loss. But this asymmetry occurs much more generally in strategic interactions. Each decision maker reacts to her and other's potential for errors even if they do not exactly know the exact error in each circumstance. One is essentially looking for a fixed point in the responses of all strategic players where each of them finds the best response to the reaction of others. This is a Nash equilibrium in a game with imperfect information.

In this paper we examine how asymmetric learning can be used in a strategic setting. We consider a simple strategic pricing game in which a seller is providing a single good that is valued differently by customers [4]. The seller does not know the class of the customer and the customer only knows her class imperfectly. In particular, the customer utilizes a classification model to predict her class and then make a purchase decision. This classification is not perfect and we examine the effect of errors on the seller, the consumer and a regulator who might be interested in maximizing social welfare.

The paper by Lewis and Sappington [5] is the closest antecedent to this work. They only consider symmetric learning in which each error type is affected equally. For the case when the market is not covered, i.e., only the high type are served, they find that the errors are always costly to the monopolist, i.e, the monopolist wants to provide as much information as possible to reduce errors. In contrast, when asymmetric learning is possible, and we separate out the effects of FN and FP errors, we find the effect of the errors on monopolist's profit depends on the type: False Positive errors increase the profit while False Negative reduce it. Johnson and Myatt [6] generalize the results in [5] but they also stay with symmetric information, considering only symmetric rotations of demand curves. They find a similar corner result as [5]. But, this too does not hold when the machine learning can be asymmetric leading to differential FP and FN error rates.

Next, we consider a model in which the consumer has a choice of two classifiers, and can use them singly or randomize between them. We show that a consumer may choose to randomize between one classifier and another that is inferior in every error rate. Conventional data science literature assumes that inferior classifiers have no utility. They typically discard inferior classifiers, by only considering classifiers that lie on the upper surface of the convex hulls of ROC curves [7].

2 The Model

Consider a monopoly seller of a good that is valued differently by different customers. Suppose that the utility from the purchase is $\theta - p$ where θ is a customer specific value of the good and p is the price set by the monopolist. Assume that the alternative of no purchase results in 0 net value to the customer. The customers are heterogeneous. A λ proportion of them have a high value θ_H for the good, while $1-\lambda$ proportion have a low value θ_L, with $0 < \theta_L < \theta_H$. We label customer with $\theta = \theta_H$ as belonging the High class and the others as belong to the Low class. For purposes of definition of error types, we label the High class as the positive class.

The monopolist does not know which customer belongs to which class, i.e., and so cannot discriminate between the customers, and consequently can only offer the good at a single price.

We consider two situations based on what the customers know about their value for the good:

1. Case p (perfect knowledge): The customers privately and perfectly know the value of the good (No Error).

2. Case e (with error): The customers have access to a classifier that give them a prediction of their class, but this prediction is not perfect.

Case p: No Errors - Customers know their type perfectly

Consider the case of perfect information: The customers know their actual class, i.e, they know whether they have a value of θ_H or θ_L. The monopolist sets a price p and customers make their purchase decision:

$p \leq \theta_L$	All customers purchase
$\theta_L < p \leq \theta_H$	Only the high type customers purchase
$\theta_H < p$	No customers purchase

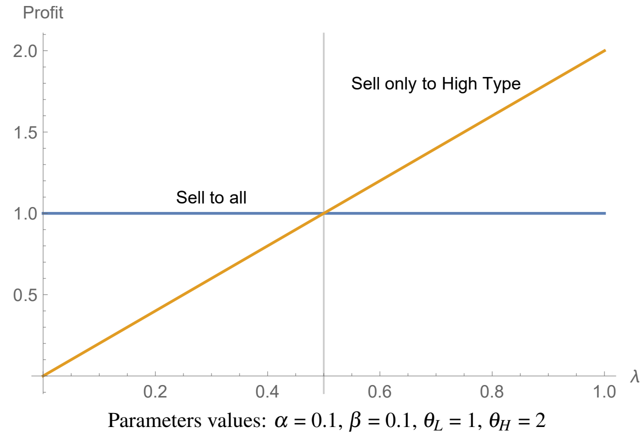


Figure 1: Profit from serving only the high type or all customers (no error)

It is easy to see that the monopolist's optimal decision is to set $p = \theta_L$ if $\lambda \leq \theta_L/\theta_H$ and $p = \theta_H$, otherwise. The graph of the profits from the two strategies is in Figure 1. They are both linear in λ with a single crossing at $\lambda = \theta_L/\theta_H$.

The expected profit, consumer surplus and the total social welfare are exhibited below:

$$pr_p = \begin{cases} \theta_L & \text{if } \lambda \leq \theta_L/\theta_H \\ \lambda\theta_H & \text{otherwise} \end{cases}$$

$$cs_p = \begin{cases} \lambda(\theta_H - \theta_L) & \text{if } \lambda \leq \theta_L/\theta_H \\ 0 & \text{otherwise} \end{cases}$$

$$sw_p = \begin{cases} (1-\lambda)\theta_L + \lambda\theta_H & \text{if } \lambda \leq \theta_L/\theta_H \\ \lambda\theta_H & \text{otherwise} \end{cases}$$

Case e: Errors in Classification by Customers

The customers have access to a classifier that give them a prediction of their class, but this prediction is not perfect. Let H and L refer to the actual customer type that is not known to customers or the monopolist. Let h and l be the predicted class. Using this notation and taking the high class to be the positive one, we can define the probabilities of error:

$$P[h|L] = \alpha \qquad \text{(False-Positive Rate or Type I Error Rate)}$$

$$P[l|H] = \beta \qquad \text{(False-Negative Rate or Type II Error Rate)}$$

Without loss of generality we assume that $\alpha + \beta \leq 1$ for if this were not the case then we can just reverse the polarity of the classification to get this inequality. We also assume that $\alpha < 0.5$ and $\beta < 0.5$.

The expected value of the product for customers depends on the signal they have received:

$$\theta_l = \theta_L Prob[L|l] + \theta_H Prob[H|l] = \theta_L \frac{(1-\lambda)(1-\alpha)}{(1-\lambda)(1-\alpha)+\lambda\beta} + \theta_H \frac{\lambda\beta}{(1-\lambda)(1-\alpha)+\lambda\beta} \tag{1}$$

$$\theta_h = \theta_L Prob[L|h] + \theta_H Prob[H|h] = \theta_L \frac{(1-\lambda)\alpha}{(1-\lambda)\alpha+\lambda(1-\beta)} + \theta_H \frac{\lambda(1-\beta)}{(1-\lambda)\alpha+\lambda(1-\beta)} \tag{2}$$

With the assumptions about α and β above, $\theta_l < \theta_h$. The customers make a decision based on the expected value after the signal and will only purchase the good if this is higher than the price. The monopolist has the choice of setting the price at θ_l and serving everyone or setting the price at θ_h and serving only those who had a high signal. The probability of getting a high signal is $p_h = \alpha(1-\lambda) + (1-\beta)\lambda$. Consequently, the monopolist's optimal decision is to set $p = \theta_l$ if $p_h \leq \theta_l/\theta_h$ and $p = \theta_h$, otherwise.

The equation determining the boundary conditions when all or only the high predicted type are served, $p_h = \theta_l/\theta_h$ is an implicit equation that is a quadratic in λ. In the next theorem we characterize the roots to this equation.

Theorem 1. *The equation $p_h(\lambda) = \theta_l(\lambda)/\theta_h(\lambda)$ has either zero or two roots in the interval $[0,1]$. In the former case, the monopolist serves all customers regardless of λ. In the latter case, the monopolist serves all customers only if the λ does not lie between the two roots, and serves only those predicted high, otherwise.*

The case when there are two roots is illustrated in Figure 2.

The expected profit, consumer surplus and the total social welfare are derived in the proposition below:

Theorem 2. *The profit for the monopolist, the consumer surplus and the social welfare in the strategic*

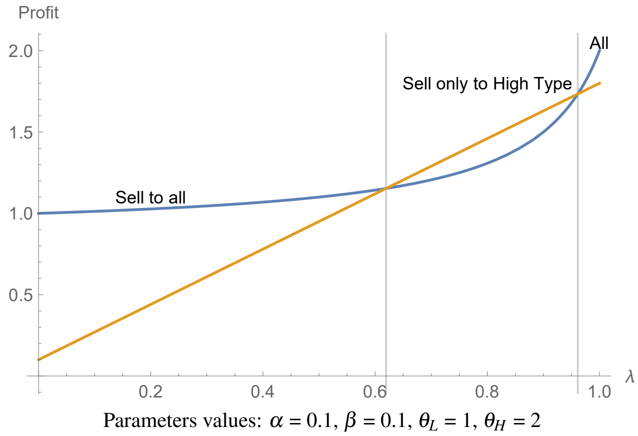


Figure 2: Profit from serving only the high type or all customers (with error)

screening game with classification errors are:

$$pr = \begin{cases} \theta_l & \textit{if all customers are served} \\ (1-\lambda)\alpha\theta_L + \lambda(1-\beta)\theta_H & \textit{if only customers predicted high are served} \end{cases} \tag{3}$$

$$cs = \begin{cases} (1-\lambda)(\theta_L - \theta_l) + \lambda(\theta_H - \theta_l) & \textit{if all customers are served} \\ 0 & \textit{if only customers predicted high are served} \end{cases} \tag{4}$$

$$sw = \begin{cases} (1-\lambda)\theta_L + \lambda\theta_H & \textit{if all customers are served} \\ (1-\lambda)\alpha\theta_L + \lambda(1-\beta)\theta_H & \textit{if only customers predicted high are served} \end{cases} \tag{5}$$

As you can see in Theorem 2, the profit, consumer surplus, and social welfare depends on whether all customers are served or only those predicted high type are served. This in turn is determined by Theorem 1.

3 The Effect of Errors on the Surplus

Let us start by looking at the cost of error on the consumers since they are the ones employing the classification mechanism to predict their type. Since our goal is to determine the value of incrementally reducing error rates using machine learning, we look at the marginal change in surplus for the entity in question. Hence for the consumers, the marginal effect of Type I error is defined is $d\ cs/d\alpha$ and the marginal effect of Type II

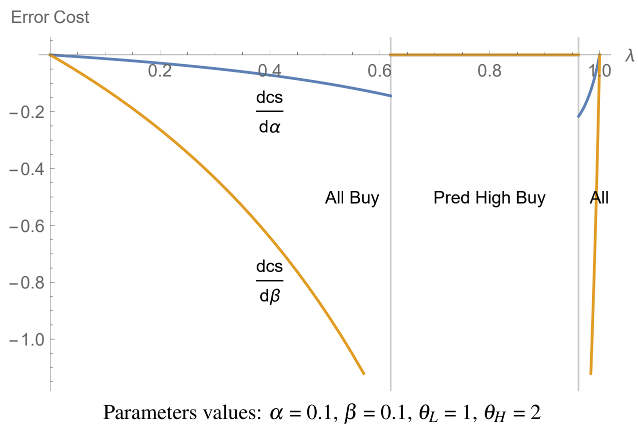


Figure 3: The Marginal Effects of Errors on the Consumer

error is defined is $d\ cs/d\beta$ where cs is the consumer surplus exhibited in Theorem 2.

In Figure 3 the cost of False-Positive and False-Negative errors are plotted against λ. Vertical grid lines at 0.62 and 0.96 indicates the values for λ where the seller is indifferent between selling to all customers or only those who got the high prediction. In particular, for $\lambda \leq 0.62$ and $\lambda \geq 0.96$, the monopolist sells to everyone. The monopolist sells to only those predicted high when the λ value lies between the two roots.

Proposition 3. *Errors of both types are costly for the consumer when the market is covered. Further, in this case the error costs are asymmetric, with the cost of False-Negative being greater than that of False-Positive. Errors have no effect on the expected consumer surplus when only the high types are served.*

While the consumers are the ones performing the prediction, the monopolist can affect the quality of the prediction indirectly by affecting the amount of product specific information available to the consumer. Thus, it is instructive to look at the incentives of the monopolist to provide information to the consumer. The results in this subsection are similar to those in [5].

As you can see in Figure 4, the only case in which the monopolist wants to help the consumer is when only the high types are served and even in that case, it only cares to reduce the incidence of False-Negative errors. In our model, a False-Negative occurs when a high type consumer wrongly predicts that she is of the low value type. Clearly, this reduces the market for the high type good as in this equilibrium, only customers who are predicted high are purchasing the product. Further, the price set at θ_h also reduces with greater incidence of False-Negatives.

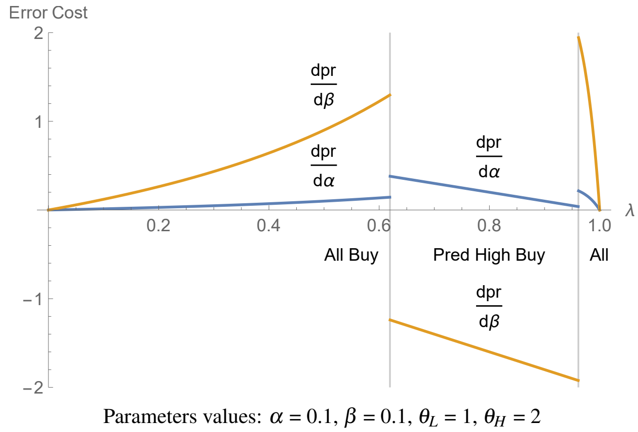


Figure 4: The Marginal Effect of Errors on the Monopolist

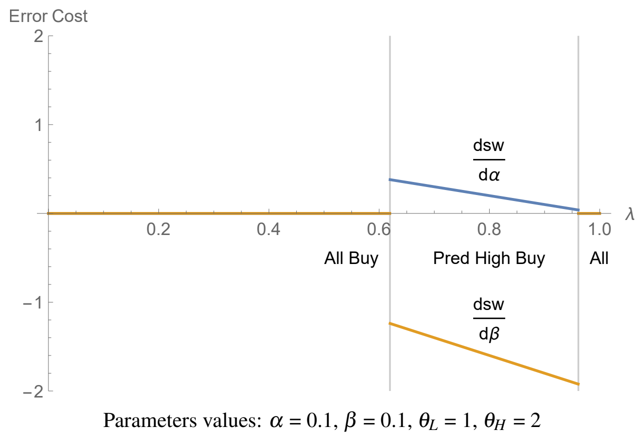


Figure 5: The Marginal Effect of Errors on the Regulator

Proposition 4. *The monopolist's profit increases with error rates in all cases except for False-Negatives when the market is not covered. In other words, the monopolist has no incentive to help improve the prediction model used by the consumers when the market is covered, and only cares about False-Negatives when the market is not covered.*

Now we turn to the effect of errors on social welfare. We assume that a regulator, if present, has an interest in increasing social welfare. In this case, the transfers between the consumer and the monopolist do not matter, except that it may change the size of the market. Consequently, we find that the regulator only cares about errors when the market is not covered. This is evident in Figure 5.

Proposition 5. *The social welfare is unaffected by classification errors when all customers are served. It increases with the False-Positive rate and decreases with the False-Negative rate when only the predicted*

	Monopolist	Consumer	Regulator
Marginal Effect of False-Positive	Positive	Negative	0
Marginal Effect of False-Negative	More Positive	More Negative	0

Table 1: Marginal Effect of Error With Fully Covered Market

high customers are served.

Let us tie together the incentives of all the parties and compare who has an incentive to reduce errors. Let us first consider the case when the market is covered. This is summarized in Table 1. The errors are costly for the consumer but profit enhancing for the monopolist. So, the customer has an incentive to improve the quality of the prediction but the monopolist is not incentivized to do so. The monopolist has an incentive to hide information and not help the consumers. But, consumers benefit from more information (reduced errors). In this situation we expect the social media and third party ranking sites to play a big role in the customer decision. Social welfare is unaffected by these errors as the market remains covered and only the price is affected.

We also note that false negative errors, in which a high type customer mistakenly classifies herself as the low type is particularly costly for the consumer because the effect it has on the price set by the monopolist. With these errors, the expected value of the product is higher (incorrectly), and this results in a transfer of surplus from the consumer to the monopolist. Do note that in this 'pooling' equilibrium, all consumers continue to purchase despite the error, it is the price that is adversely affected. Since the effect of false negatives is much larger than that of false positives, we can expect the consumer to tailor their machine learning process to favor reduction of false negatives. As we discussed in the introduction, they can do this by re-sampling, altering the cutoff score, or giving observations with low type more weight.

Now consider the case when only the customers who are predicted high are served. The costs are summarized in Table 2.

In this case we find that the different error types have different impact on profit: While false negative errors are costly for the monopolist, false positive errors increase the profit. This is directly related to the pool of customers who are predicted high, the only customers who purchase the product in this case. False

	Monopolist	**Consumer**	**Regulator**
Marginal Effect of False-Positive	Positive	0	Positive
Marginal Effect of False-Negative	Negative	0	Negative

Table 2: Marginal Effect of Error With Only Those Predicted High Served

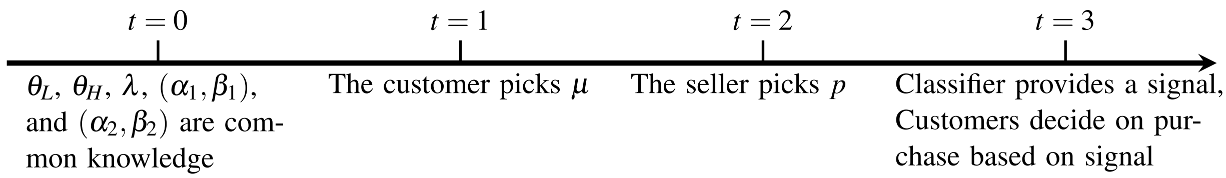


Figure 6: Timeline of Classifier Choice

positive errors increase the size of the pool while false negatives reduce the size of the pool. The expected customer welfare is unaffected as the price is set at the expected value of the product. So, the consumer has no incentive to invest in better information, while the monopolist has an incentive to make more customers perceive the good to be of high quality. The monopolist may take unilateral actions such as advertising in this case.

The regulator also has a strong interest in reducing false-negatives as these customer who would have bought the product do not do so because of errors in prediction. The cost of errors is asymmetric, in that while the false negatives are costly, the false positives are not. So if the regulator facilitates any improvement in learning, such as by setting up information sites, it would focus its approach on improving learning in the positive class so as to reduce false negatives.

4 Classifier Choice by Consumer

In the previous section, we took the classifier and its error rates as a given and explored the incentives for each party to adjust the rates on the margin. We now extend the model by considering classifier choice by the consumer. We consider a simple model in which the customer can use any of two classifiers or randomize between them. These classifiers and their error rates are public knowledge. Figure 6 provides a time line of the decision and the information available at each stage.

At time 0, the two classifiers are given by specifying their error rates: (α_1, β_1), and (α_2, β_2).

At time 1, the customer picks $\mu \in [0,1]$ for randomizing between the two classifiers. This produces a classifier with errors: $(\alpha = (1-\mu)\alpha_1 + \mu\alpha_2, \beta = (1-\mu)\beta_1 + \mu\beta_2)$.

At time 2, the seller picks the price p and at time 3, the customers see the signal from the classifier and make expected utility maximizing purchases conditional on the signal seen.

In, the prior section we have already worked out the decisions at times 3 and 2. Now let us turn to time 1. To illustrate the choice of a classifier, let us consider two cases:

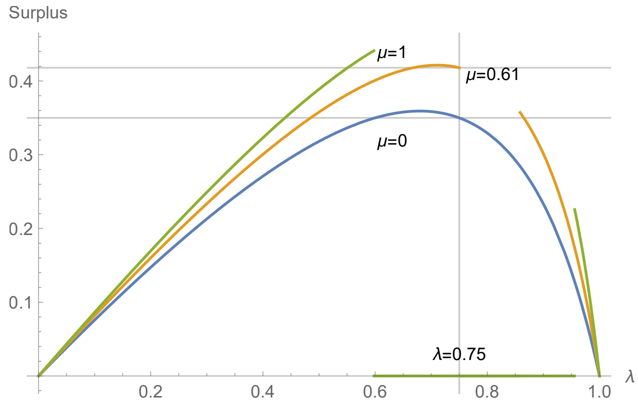
Case 1: Non-dominant Classifiers	$(\alpha_1 = 0.1, \beta_1 = 0.2),\ (\alpha_2 = 0.2, \beta_2 = 0.1)$
Case 2: Dominating Classifier	$(\alpha_1 = 0.1, \beta_1 = 0.1),\ (\alpha_2 = 0.2, \beta_2 = 0.2)$

4.0.1 Non-dominating Classifiers

In this case, classifer 1 has a lower FPR (α), but a higher FNR (β) than classifier 2. Neither classifier dominates the other in error rates. We will show that given a choice of randomizing between them, the customer strictly prefers a classifer that randomizes between these. In Figure 7 the consumer surplus is plotted against λ. Consider the case when $\lambda = 0.75$. In this case, if the customers set $\mu = 0$, i.e., they pick classifier 1 which has errors $(0.1, 0.2)$ then all customers, whether they see the low or the high signal will buy and their consumer surplus will be 0.35. Now if the customers set $\mu = 1$, i.e., they all pick classifier 2 which has errors $(0.2, 0.1)$ then only the high signal will buy and their consumer surplus is zero. Clearly, they get a higher surplus with $\mu = 0$ than with $\mu = 1$. But, as the figure shows, the customers can do even better by picking $\mu = 0.61$ for with that choice, all customers again buy but get a surplus of 0.42 that is higher than what is obtained with $\mu = 0$.

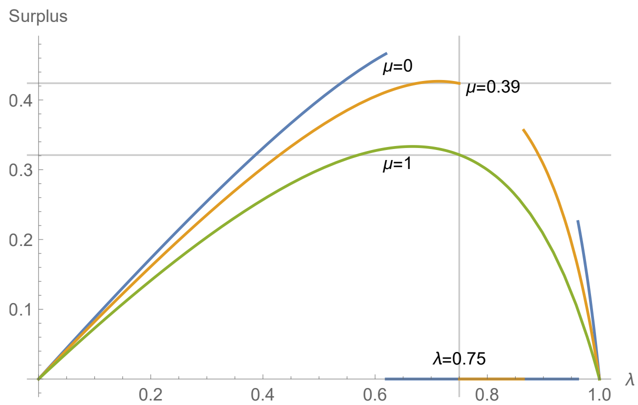
4.0.2 Dominating Classifier

In this case, classifier 1 has lower FPR and FNR than classifier 2. This situation is illustrated by Figure 8. Once again, consider the case when $\lambda = 0.75$. If the customers set $\mu = 0$, i.e., they pick classifier 1 which



Parameters values: $\alpha_1 = 0.1$, $\beta_1 = 0.2$, $\alpha_1 = 0.2$, $\beta_1 = 0.1$, $\theta_L = 1$, $\theta_H = 2$

Figure 7: Classifier Choice by Customer in Non-dominating Case

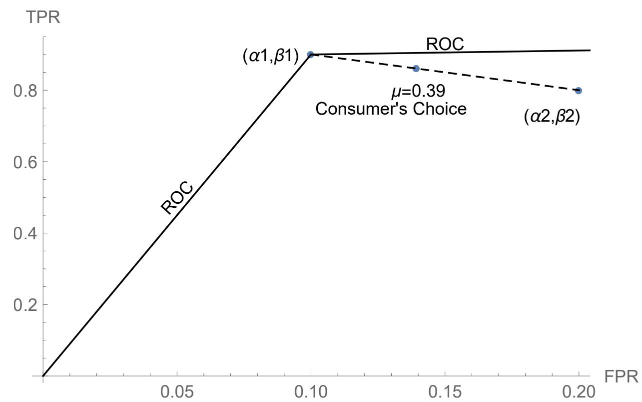


Parameters values: $\alpha_1 = 0.1$, $\beta_1 = 0.1$, $\alpha_1 = 0.2$, $\beta_1 = 0.2$, $\theta_L = 1$, $\theta_H = 2$

Figure 8: Classifier Choice by Customer in the Dominating Classifier Case

has errors $(0.1, 0.1)$, the better of the two classifiers, then only the high type customers will buy and the consumer surplus is 0. However, if they set $\mu = 1$, i.e., they pick the inferior classifier with errors $(0.2, 0.2)$, they em get a higher surplus of 0.32 as all customers purchase in this case. In fact, the customers can do even better. If the customers set $\mu = 0.39$ then get a even higher surplus of 0.42.

Consider Figure 9 which shows the ROC curve. With just these two classifiers, only classifier 1 is on the rROC curve. Classifier 2 in below the ROC curve and showed above that the customer prefers classifier 2 to classifier 1. The best choice is to set $\mu = 0.61)$ and this classifier lies on the line between classifier 1 and 2. This result shows that strategic considerations overturn the conventional data science literature indicates that classifier 2 will never be utilized by the customers as it has performs poorly in all respects [7].



Parameters values: $\alpha_1 = 0.1$, $\beta_1 = 0.1$, $\alpha_1 = 0.2$, $\beta_1 = 0.2$, $\theta_L = 1$, $\theta_H = 2$

Figure 9: ROC Curve and Customer's Classifier Choice

References

[1] C. Elkan, "The foundations of cost-sensitive learning," in *IJCAI*, 2001.

[2] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358 – 3378, 2007.

[3] K. M. Ting, "A comparative study of cost-sensitive boosting algorithms," in *In Proceedings of the 17th International Conference on Machine Learning*. Morgan Kaufmann, 2000, pp. 983–990.

[4] M. Musa and S. Rosen, "Monopoly and product quality," *Journal of Economic Theory*, vol. 18, pp. 301–317, 1978.

[5] T. Lewis and D. Sappington, "Supplying information to facilitate price discrimination," *International Economic Review*, vol. 35, no. 2, pp. 309–327, 1994.

[6] J. P. Johnson and D. P. Myatt, "On the simple economics of advertising, marketing, and product design," *The American Economic Review*, vol. 96, no. 3, pp. 756 – 784, 2006.

[7] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Machine Learning*, vol. 42, no. 3, pp. 203 – 231, 2001.

Appendix

Theorem 1 The equation $p_h(\lambda) = \theta_l(\lambda)/\theta_h(\lambda)$ has either zero or two roots in the interval $[0,1]$. In the former case, the monopolist serves all customers regardless of λ. In the latter case, the monopolist serves all customers only if the lambda does not lie between the two roots.

Proof. The proof of the theorem follows from the following: (1) $p_h\theta_h$ is linear. (2) θ_l is convex increasing in λ with the values at $\lambda = 0$ and $\lambda = 1$ above $p_h\theta_h$. □

Theorem 2: The profit for the monopolist, the consumer surplus and the social welfare in the strategic screening game with classification errors are:

$$pr = \begin{cases} \theta_l & \text{if all customers are served} \\ (1-\lambda)\alpha\theta_L + \lambda(1-\beta)\theta_H & \text{if only customers predicted high are served} \end{cases}$$

$$cs = \begin{cases} (1-\lambda)(\theta_L - \theta_l) + \lambda(\theta_H - \theta_l) & \text{if all customers are served} \\ 0 & \text{if only customers predicted high are served} \end{cases}$$

$$sw = \begin{cases} (1-\lambda)\theta_L + \lambda\theta_H & \text{if all customers are served} \\ (1-\lambda)\alpha\theta_L + \lambda(1-\beta)\theta_H & \text{if only customers predicted high are served} \end{cases}$$

Proof. First consider the case when the market is covered. As discussed before, in this case the monopolist sets the price as θ_l and all customers buy the product. In this case, the expected consumer surplus is $(1-\lambda)(\theta_L - \theta_l) + \lambda(\theta_H - \theta_l)$ and the social surplus is $(1-\lambda)\alpha\theta_L + \lambda(1-\beta)\theta_H$.

Now consider the case when only the high type are served. In this case the monopolist sets the price at θ_h and sells only to the customers predicted high, i.e., p_h of them. Substituting from equation **??**, the profit is $\theta_h p_h = (1-\lambda)\alpha\theta_L + \lambda(1-\beta)\theta_H$. The consumer surplus is $(1-\lambda)\alpha(\theta_L - \theta_h) + \lambda(1-\beta)(\theta_H - \theta_h) = 0$. Since the consumer surplus is zero, the social welfare is equal to the profit.

□

Proposition 3 Errors of both types are costly for the consumer when the market is covered. Further, in this case the error costs are asymmetric, with the cost of Type II Error being greater than that of False-Positive

Error. Errors have no effect on the expected consumer surplus when only the high types are served.

Proof. When the market is covered,

$$\frac{\partial cs_e}{\partial \alpha} = \frac{-\beta(\theta_H - \theta_L)\lambda(1-\lambda)}{(1-\alpha-(1-\alpha-\beta)\lambda)^2}$$
$$\frac{\partial cs_e}{\partial \beta} = \frac{-(1-\alpha)(\theta_H - \theta_L)\lambda(1-\lambda)}{(1-\alpha-(1-\alpha-\beta)\lambda)^2}$$

Since we have assumed that $\theta_H > \theta_L$ and that $\alpha + \beta < 1$ without loss of generality, the result follows. □

Proposition 4: The monopolist's profit increases with error rates in all cases except for False-Negative error when the market is not covered. In other words, the monopolist has no incentive to help improve the prediction model used by the consumers when the market is covered, and only cares about False-Negative error when the market is not covered.

Proof. First consider the market covered case. In this case the profit = θ_l. The derivatives of θ_l with respect to the error rates for the market covered case is shown below.

$$\frac{\partial \theta_l}{\partial \alpha} = \frac{\beta(1-\lambda)\lambda(\theta\mathrm{H} - \theta\mathrm{L})}{(\lambda(\alpha+\beta-1)-\alpha+1)^2}$$
$$\frac{\partial \theta_l}{\partial \beta} = \frac{(1-\alpha)(1-\lambda)\lambda(\theta\mathrm{H} - \theta\mathrm{L})}{(\lambda(\alpha+\beta-1)-\alpha+1)^2}$$

It can be readily seen that both the derivatives are positive and hence the cost of errors are negative.

Now consider the case when the market is not covered. In this the profit is $ph\theta_h$.

$$\frac{\partial ph\theta_h}{\partial \alpha} = (1-\lambda)\theta_L$$
$$\frac{\partial ph\theta_h}{\partial \beta} = -\lambda\theta_H$$

It can be readily seen that the derivative wrt to α is positive while that wrt β is negative. Hence, the costs are negative, and positive, respectively. □

Proposition 5 The social welfare is unaffected by classification errors when all customers are served. It increases with False-Positive error rate and decreases with False-Negative error rate when the only the predicted high customers are served.

Proof.

$$\frac{\delta sw_e}{\delta \alpha} = (1-\lambda)\theta_L$$
$$\frac{\delta sw_e}{\delta \beta} = -\lambda \theta_H$$

The proposition readily follows from the equations above. □