Insurance Pricing Competition

Guidance document

Introduction

Insurance pricing is about predicting risk

Car insurance prices have historically been calculated based on how "risky" a driver is. This measure of risk used to be computed based on only a few characteristics. Today, most insurance companies (97% in fact) use Generalised Linear Models to produce an insurance premium quote for car insurance. However, the industry is slowly starting to use more advanced machine learning techniques such as Decision Trees, Neural Networks, and the like. No one yet knows which one is more practical for an insurance company. Hopefully you will find out!

Your challenge is to make money by predicting the correct risk

In this competition, you play the role of a new insurance company, and with some data, must create a pricing model to compete for price-sensitive customers and make money!

Some customers will have a car accident and thus, as their insurer, will cost you money. You must balance offering cheap prices to attract customers with making sure you have enough money to pay off all the potential insurance claims.

The top three models win Amazon vouchers (Deadline January 12 2020)

Now that you are signed up and have the data, you have until **January 12 2020** to submit your model via our submission platform (see below). We will then create a market where all models compete with each other and the top three most profitable firms will be awarded with Amazon vouchers of £50, £30, and £20 respectively. That and of course:

ETERNAL GLORY.

Insurance pricing in a nutshell

A typical pricing model has two parts

Insurance pricing models are responsible for figuring who is a risky customer and pricing their insurance accordingly. Typically, they are split into two components: a frequency model (F) and a severity model (S). The first part predicts the probability of having an accident, while the second predicts the expected cost of an accident. The frequency model is typically a classifier, predicting whether a contract will default, from which the probabilities are extracted, whereas the severity model is a regressor. The final pure insurance premium (P) is the product of the two:

$$P = F \cdot S$$

You are allowed to ignore severity if you choose to

The severity model is typically quite challenging. Therefore, you are free to only submit a frequency model (F) and multiply by the average price. To evaluate whether this is a good idea, you can use the AI market that we have provided before you submit your model (see below).

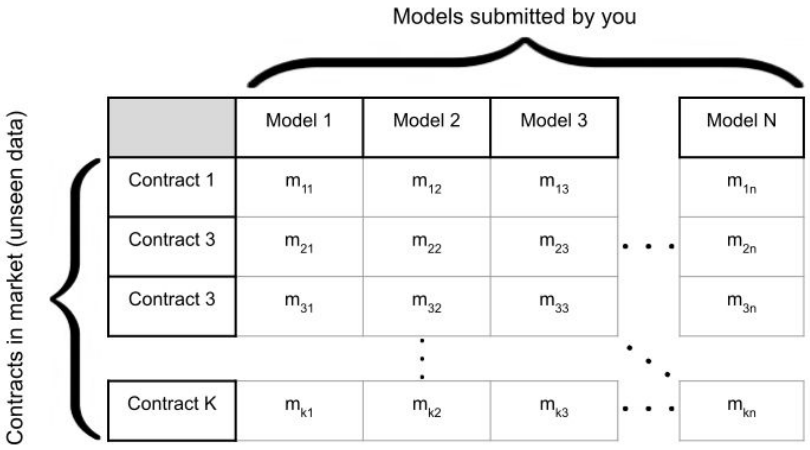
Competition structure

You have to pick between two model types: Linear or Otherwise

Just like a real company, in this competition, you must also choose between either a linear model (like most of the industry today) or something else. You are also allowed to submit both types.

Customers always go for the cheapest price offered to them

Each customer i, is offered a series of prices (m_{ij}) by each player j, (competition participants). They then pick the cheapest price offered to them and that particular player wins that contract. Note that if your prices are too expensive then you might not get any contracts!



You can use the AI market to adjust your pricing strategy

Before you engage in competition with other players, you can use the AI market to see how competition impacts your model. This is a market pre-populated with some standard pricing models that should not be too hard to beat. Once you submit a model, it is automatically put in this market. Some aggregate statistics are provided to you such as profit, market share, average price and so on.

Coding environment, data, model submission

The training data and the data dictionary are available online

The training data for this competition is made up of 80,000 real car insurance contracts from the mid 2000s. The data is from a major French car insurer and has been enriched with geographic information. You are not allowed to use any external data in addition to what has been provided. The training data is available on dropbox bit.ly/PricingGame2019 with the password **icdssinsurancepricing2019**.

To participate, sign up on the submission platform pricing-game.dsi.ic.ac.uk

We have built pricing-game.dsi.ic.ac.uk for model submission. Go ahead and sign-up, it should take less than 2 minutes! The contest is named **Pricing Game (ICDSS)**, and the password is **boatapplefoxmath**.

You can use Python or R, using the most commonly used packages

Our platform supports both Python and R. We will provide you with code templates for each of these that you can fill. Further, we will provide you with a list of supported packages. Both the templates and the supported packages are on the same dropbox bit.ly/PricingGame2019. For R, if you want to install a package in your submitted code, it must be from CRAN (the official R archive).

Privacy

You will all be removed from our mailing list by the end of the competition, which is when we provide you with the results and distribute the prize! Further we will use the models you use for research.

If you wish to be removed before this point, please email pricing-game@imperial.ac.uk with the subject "UNSUBSCRIBE". You can also email us with any questions!