Pricing your Pooch

EE 500 Fall 2017

Kody G. Ferguson | kgfergus@usc.edu

Rezwan A. Rasul | rrasul@usc.edu

The dog breeding industry defies normal pricing models and conventional consumer behavior. When selling a living animal, especially with the intention of companionship, breeders want to ensure that it will go to a home that can adequately take care of it and give it the environment it needs to thrive. As such, a relationship must be established between the breeder and buyer prior to the purchase. This ensures a mutual approval of the transaction so both parties are satisfied. The atypical nature of selling a dog extends to their pricing. By speaking and meeting with over 15 dog breeders around the US, we found that the most influential factors when pricing a new puppy is its breed and pedigree. By pedigree, we are referring to the champion-status of its parents and even the champion-status of ancestry further up its genetic tree. No government or corporate entity oversees the sale of dogs and as such no consistent pricing model exists. Using 3,100 data points collected from breeders and online listing platforms, we have created a model that prices new, purebred pups given a set of 12 features.

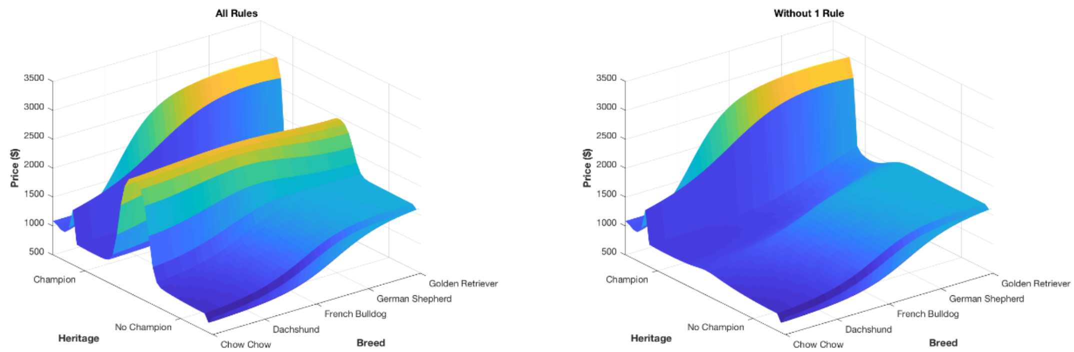
The feature space for each dog breed consists of 12 features, of which 4 are categorical (i.e. the breed, heritage) and the other 8 are numerical and binary (i.e. vaccination status, health guarantee, etc.). First, we employed a neural network to reduce the feature space. Then, with the reduced feature space we trained a Gaussian ASAM with 30 rules to create a reliable pricing model.

We modified the neural network (DNN) regression model provided in [1] [2]. For evaluation purposes, we used the average loss as a function of training epoch and the normalized root mean square (rms) error of the test set. We found that dividing the dataset into an 80:20 train/test ratio provided the lowest rms error compared to other tested ratios. We chose a batch size of 500 because it provided a smoother decrease of average loss while not increasing the final rms error, compared to smaller batch sizes. When we vary the network parameters, (such as number of hidden layers, number of neurons, optimizer type, etc.) we did not observe any significant change in the regression performance. Removing features, on the other hand, increased the final rms error as expected. Interestingly, the impact varies according to the features removed. We concluded from the simulation that the most important factors to determine a puppy's price are its breed and lineage, which reaffirmed the expert's opinion that we had collected. After optimization, the normalized test-set rms error is 0.78σ, or equivalent to \$510 in price, when utilizing the total feature space.

Using the experimental observation that breed and lineage contribute most to a puppy's price, we created a two-dimensional fuzzy system whose rule count, standardization, and noise parameters had been optimized through automated tests. When increasing the rule count, we found that the rms error for the test set converged after ~30 rules. The model achieved an rms error less than 0.8σ with only 2 features, a value close to NN for >15 rules. We also observed that the impact of removing rules after training with a specific number of rules is more adverse than training the network with less fuzzy rules to begin with. Next, we simulated the fuzzy system and approximated the price for 20 different dog breeds. Simulation confirms that a dog with champion lineage tends to be more expensive compared to the ones with general lineage. The fuzzy system also suggests that the most expensive breed is the

[1] https://github.com/tensorflow/tensorflow/blob/master/tensorflow/examples/learn/boston.py py
[2] https://github.com/tensorflow/models/tree/master/official/wide_deep/

French bulldog, with its price almost solely decided by its breed. We took out one specific rule that patched the 'bump' of this specific breed and observed a sharp decline of predicted price by $820.



This project allowed us to explore an ambiguous problem through the power of neural networks and fuzzy systems. We were fortunate to be able to confirm our expert knowledge with experimental results as well as gain new insights into dog pricing. Looking forward, more accurate models could be made for each breed by using detailed descriptors of champion heritage rather than a simple binary feature. A larger data set and specific region of sale may also garner a deeper insight into dog pricing patterns. Dog pricing may continue to be a fuzzy problem in the coming years, but with the rise of machine learning, we hope many of these data driven findings can be shared with breeders to improve their business and livelihood.

[1] https://github.com/tensorflow/tensorflow/blob/master/tensorflow/examples/learn/boston.py py
[2] https://github.com/tensorflow/models/tree/master/official/wide_deep/