

1. [LFD Exercise 1.2] Suppose that we use a perceptron to detect spam messages. Let's say that each email messages represented by the frequency of occurrence of keywords, and the output is +1 if the message is considered spam.

1. Can you think of some keywords that will end up with a large positive weight into perceptron?

2. How about keywords that will get a negative weight?

3. What parameter in the perceptron directly affects how many borderline messages end up classified as spam?

Ans:

Keywords that lead to a positive weight indicates that messages with the keyword are more likely to be spamming. Some words that are very casual or oral, or leading receiver to pay money are likely to be spamming. For examples, 'pay', 'credit card' or 'buy'.

Keywords contribute a negative weight if they are not likely to be spamming. If words are very written or official, they are likely to be not spamming. Examples like 'purchase'. Parameters to control the strictness of the checking are like message length or message sending rate.

2. Consider a coin tossing experiment. You toss a coin 100 times, with the result of heads 70 times and tails 30 times. We denote the probability of heads of this coin as Θ . Now consider a coin toss.

1. Build a model using maximum likelihood estimation (MLE) to infer Θ .

2. Can we judge that this is an unfair coin? Explain your answer.

We can probably use logistic regression to do the estimation of classification. If head faces up, note the result as 1, otherwise 0. A sample model is to use sigmoid function $= 1 / (1 + e^{-f(x)})$ in which $f(x)$ is a linear regression with regards to the parameters like Throwing angle, starting speed, or air speed. Suppose the measurements of the parameters are possible, we can conduct logistic regression to gain the weightage for each attribute.

We cannot conclude that this is an unfair coin, since we need to analyze the obtained model. The above mentioned attributes are likely to affect the final result. The higher chance of head tossing may be due to the person's throwing sense. We will need to analyze the obtained weightage and compare it with a normal coin to see whether this one is unfair or not.

3. In the programming logistic regression, part (c), we did away with the stochastic idea of SGD and substituted a round-robin version, which deterministically uses the next point in turn to perform the gradient descent. Describe whether you think this is a good robust idea or not for datasets in general.

A round robin method may not be robust since it may be stuck in the local extreme value. If the function has both minimum and global minimum, a round robin method is likely to stop at a local minimum, which is not really optimal. There must be some methods to escape from local extreme value. SGD method can avoid local minimum by randomly jump through the local extreme value. Besides, since a deterministic method is used, the obtained result must be same every time to run it. This is not suitable to experiment for multiple different results.