A review so far
through previous exam question

May 28, 2018

Question 1.

- [ ] The cosine similarity of two unit vectors (vectors of length 1) is their dot product.
- [ ] The cosine similarity of two vectors is 0 if the vectors are orthogonal.
- [ ] If two random variables depend on each other, their correlation coefficient is always positive.
- [ ] Introducing non-linear basis functions causes logistic regression error function to be non-convex.
- [ ] Dot product of two vectors is another vector.
- [ ] Covariance of two variables is not affected by the units of measurement of each variable.
- [ ] The expected value (mean) of a random variable is its most likely value.
- [ ] The output of a probability density function has to be in the interval $[0, 1]$ (between 0 and 1, inclusive).
- [ ] If the mutual information between two random variables is 1, we can predict one from the other with certainty.
- [ ] KL-divergence of two distributions P and Q, $D_{KL}(P \parallel Q)$, is always larger than entropy of P, $H(P)$.
- [ ] Regularized estimation of machine learning models reduces the model's fit to the training data.
- [ ] If a machine learning model has a convex loss function, one can calculate the minimum loss analytically.
- [ ] No correlation with the outcome variable is a desired property of the predictors for a statistical model.
- [ ] The perceptron algorithm adjusts the weights after every correctly classified training sample.
- [ ] The gradient of a multivariate function is the 0 vector only at the global minimum of the function.
- [ ] Multiplying a matrix with its eigenvector does not change the direction of the vector.
- [ ] A machine learning system with high recall is likely to produce few false positives.
- [ ] L2 norm of a vector in $\mathbb{R}^n$ (for any n in range $(1, \infty)$) is always smaller than or equal to its L1 norm.
- [ ] If the mutual information between two random variables x and y, $MI(x, y) = 0$, conditional entropy $H(y \mid x) = H(y)$.
- [ ] L1 regularization results in sparse parameter estimates.
- [ ] Increasing model complexity (e.g., number of parameters) in a machine learning model is likely to decrease test error.

Question 2. *Logistic regression for sentiment analysis* (15 P.)

In a binary sentiment analysis task, a logistic regression classifier is trained with the following three predictors:

x_1 number of positive emoticons, for example ':-)'

x_2 number of negative emoticons, for example ':-('

x_3 text is in 'all caps', for example 'GREAT PRODUCT'

The resulting classifier was summarized with the following equation.

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-0.1-0.5x_1+0.5x_2+0.001x_3}}$$

Based on this information, answer the following questions, and explain your answer briefly (with at most three sentences).

Assuming the emoticon predictors predict the corresponding sentiments, for which class $y = 1$?

Which class would be predicted for an input document containing lowercase text, 11 positive and 12 negative emoticons?

We have the intuition that capitalization does not indicate negative or positive sentiment, but strengthens the effects of the other features. Suggest an alternative formulation of the model (the set of predictors) that reflects this intuition.