Sample Questions-Final

March 8, 2022

A student is training a regression model to fit some data. We don't know what she is doing with her model, just that it is a regression model. After training, she notices that **the test loss is very high**. Then the student checks the training loss.

(1) If **the training loss is very low**, what would you suggest the student to do

(2) If **the training loss is also very high**, what would you suggest the student to do

---

(3) What are the differences and similarities between a step function and a sigmoid function?

---

(4) What are the differences and similarities between linear regression and linear classification?

---

(5) Which regularizer (L1 or L2) is better for feature selection? Why?

Which of the following functions are convex:

(6)
$$y = x^2 - 4$$

(7)
$$= {}^3{-}8$$

---

(8) What is the purpose of a cost function?

---

(9) What can cause erratic behavior of your cost when running gradient descent?

---

(10) Given the following confusion matrix, please calculate the true positive rate, false positive rate, and the F1 score.

	Predicted: 0	Predicted: 1
Actual: 0	900	900
Actual: 1	100	100

(11) After running gradient descent and getting back a list of weights and costs, which weights correspond to the "learned" weights of your model?

(12) When is gradient descent guaranteed to find the global minimum?

(13) How would you modify an L2 regularizer so that certain weights are regularized more than others?

(14) Can the cost of a model ever be less than 0?

0.1 New Questions

(15) Which clustering method is best when you have no knowledge about the distribution of your data and how many clusters you may expect?

(16) When would DBSCAN be preferred over K-means?

(17) When would K-means be preferred over DBSCAN?

(18) What is a good strategy for random seed selection in K-means?

(19) If computaional cost was not an issue, would you use SSE or Silhouette to evalute the quality of a clustering?

(20) When would the Silhouette score for a point be 0?

(21) When would the SSE for a clustering be 0?

(22) Which clustering method can best deal with outliers?

(23) Assume that in a very large dataset you have a few points with missing features. What should you do in this case?

(24) Assume that in a very small dataset you have a few points with missing features. What should you do in this case?

(25) What is the assumption that makes "Naive Bayes", "Naive"?

(26) If the prior for a hypothesis is very large, does that mean that you need more or less evidence to support that hypothesis?

(27) Why do you not need to calculate the marginal probabilities for Naive Bayes?

(28) What is the fusion rule used to turn binary classifiers into a multiclass classifier.

(29) How would you generate an ROC curve for a multiclass classifier?

(30) Would a very large λ in the SVM soft-margin cost lead to overfitting or underfitting?

(31) How do you measure the confidence of a classification made by a linear binary classifier?