
CS282 Machine Learning 2021 Fall

Quiz 2

1 [Lecture 9-10]

In regression, the expected mean squared error of an estimator can be decomposed in terms of bias, variance, and noise. Please interpret the above terms of expected error, bias, variance, and noise in your own words (do not copy the decomposition formula here).

2 [Lecture 11-12]

(a) In SVM, when solving for a hyperplane specified by $w^T x + b = 0$, one can always set the margin to 1, i.e., $w^T x_1 + b = -1$ and $w^T x_2 + b = 1$, where x_1 and x_2 are two support vectors and the data is separable. True or False? Explain why.

(b) Linear SVMs have no hyperparameters that need to be set by cross-validation. True or False? Explain why.

3 [Lecture 13]

Consider the data on 3 attribute $a \in \{L, M, H\}$, $b \in \{T, F\}$, and $c \in \{T, F\}$, where c is the target classification. In this question we will consider decision-tree learning based on this data.

	a	b	c
e_1	L	F	F
e_2	L	T	T
e_3	M	F	F
e_4	M	T	T
e_5	H	F	T
e_6	H	T	T

(a) What is a good attribute to split on first? Explain why.

(b) Draw a decision tree that the decision tree learning algorithm we learned in this course could build. Explain why.

4 [Lecture 14]

What is the effective number of parameters of K nearest neighbor algorithm? Explain why.

5 [Lecture 15]

In spectral clustering (graph cut), we introduce the below four types of graph cut criteria, i.e., $cut(A, B)$, $assoc(A, B)$, $Ncut(A, B)$, and $Nassoc(A, B)$. Please interpret

(a) these four criteria in your own words.

(b) the reason why we prefer $Ncut(A, B)$ and $Nassoc(A, B)$ instead of $cut(A, B)$ and $assoc(A, B)$?

6 [Lecture 16]

The columns of V in the SVD called the right singular vectors of A , and the columns of U are called the left singular vectors. Interpret the right and left singular vectors for the document term matrix.