

Quiz 3

Name:

NetID:

1. [Regression (10 points)]

- (a) Explain how linear regression can be used for nonlinear prediction from the data.

Solution: By appending each input vector with nonlinear functions of some or all of the components of the input vector.

- (b) Explain the purpose of regularization in the linear regression optimization.

Solution: Regularization allows us to deemphasize (give small or zero weights to) features that are of minor relevance to the prediction that the regression algorithm is trying to accomplish.

2. [Clustering (5 points)]

Explain how we can choose K for K -means clustering if we do not have prior knowledge of K .

Solution: We use the "knee" method, in which we plot an estimate of $J^*(K)$, obtained by running the K -means algorithm with several starting points, as a function of K and look for a sharp change in slope (knee) in the curve to choose K .

3. [Handling Data (5 points)]

Karen, a student who took ECE 365, graduated and got a job that involved machine learning. She was given a classification task, and so she trained a Naive Bayes, a Logistic Regression and an LDA classifier, and compared the training errors for the three classifiers. She found that the Naive Bayes classifier had the smallest training error, and therefore she chose to use that one for the classification task. What did Karen do wrong? What should she have done?

Solution: Karen should have used a validation step or cross-validation to compare the prediction errors on the validation data to do the model selection.

4. [Vector Quantization (10 points)]

Describe in words how clustering can be used for vector quantization based lossy compression of images.

Solution: We break the image into blocks of $n \times n$ pixels, apply K -means clustering on the blocks, and use the cluster centers and the indices of the cluster centers to represent the image. At the decoder we replace the block with the cluster center corresponding to the index of the block.