COMP-551-001 Applied Machine Learning
Self-Assessment Questions

Sarath Chandar

Lecture-01

1. In the class, we have seen that both classification and regression are supervised learning problems. What is the differece between these two supervised learning problems?
2. Give five real life applications for classification and five real life applications for regression.
3. When would you say that a particular model is a linear model?
4. What is the significance of w_0 (bias) in the linear model: $y = w_0 + w_1 x$? Why not use a model like $y = w_1 x$?

Lecture-02

1. What is overfitting? How can we find if a model is overfitting to a particular dataset?
2. Suggest at least 3 approaches to solve the overfitting problem.
3. What is the hyperparameter of a model? How is it different from the parameters of the model? How can we choose these hyperparameters?
4. While doing model selection, we choose the best hyperparameter based on the validation set performance. What will happen if we choose the best hyperparameter based on training set performance? What will happen if we choose the best hyperparameter based on the test set performance? Do we really need a separate validation set?
5. What are the hyper-parameters of the linear regression model? What are the hyper-parameters of the k-NN classifier?

Lecture-03

1. Compare and contrast least squares approach and nearest neighbor approach in terms of bias and variance.
2. In the class, we have seen that if we use squared error loss, then the expected prediction error is minimized by the conditional mean. Explain how nearest neighbor approach and least squares approach are trying to approximate this conditional mean.
3. What is Bayes rate? We have seen in the class that Bayes rate is the best possible performance any classifier can achieve. What does a classifier require in order to achieve this optimal error rate?

Lecture-04

1. What is the advantage of using non-linear basis functions with a linear model like linear regression?
2. What is the pseudo-inverse of a matrix? How is it different from the inverse of a matrix? When will the psuedo-inverse and inverse be equivalent?
3. Explain the geometrical interpretation of least squares approach.
4. When can one resort to gradient descent to minimize the objective function?
5. What happens when the step size is too large in gradient descent? What happens when the step size is too small?
6. What is the difference between gradient descent and stochastic gradient descent?
7. Gradient descent can always find the global minimum. True or False? If false, is there any scenario when it is guarenteed to find the global minumum?

Lecture-05

1. What is inductive bias? What is the inductive bias of linear regression and nearest neighbor algorithms?
2. A hypothesis which minimizes the empirical risk is also guarenteed to minimize the true risk. True or False?
3. Define bias and variance. Explain the bias-variance tradeoff.
4. Define Occam's razor.
5. Adding regularization controls overfitting. True or False?
6. Can we use L1 regularization and L2 regularization for feature selection? If so, explain how will you do that. Will there be any difference in the feature selection procedure based on whether the reuglarizer is L1 or L2 regurlarizer?
7. L1 regularization prefers sparse models. Justify.
8. Compare the geometrical views of L1 regularization and L2 regularization and argue why L1 regularizer sets more weights to zero than L2 regularizer.

Lecture-06

1. What are the three approaches to solving classification problem? Sort them in ascending order of procedure complexity.
2. Why are generative models called as *generative* models?
3. What are linearly separable problems? Give cartoon examples for linearly separable 2-class problem and not linearly separable 2-class problem.
4. In a linear discriminant model, decision surface is perpendicular to the weight vector. Prove.
5. Explain the difference between one-vs-rest classifier and one-vs-one classifier.
6. Explain various ways of solving multi-class classification problem using discriminant functions.

Lecture-07

1. Least-squares solution lacks robustness to outliers when used for classification. Justify.
2. List down the applications of PCA.
3. Why should we constrain the norm of the projection vector in PCA to 1?
4. PCA and LDA project data from one space to another space. How can we use such algorithms for classification? Which projection will be more helpful to design a classifier and why?
5. Which one of the following projection algorithms is supervised? PCA or LDA?

Lecture-08

1. What is the difference between a linear model and a generalized linear model?
2. In GDA, covariance matrix of all the class conditional densities are shared. How is this affecting the decision boundary?
3. Explain the i.i.d assumption.
4. What is the difference between GDA and QDA?
5. Define confusion matrix. How will an ideal matrix look like?
6. Define precision and recall. List two applications where precision is more important and two applications where recall is more important.
7. Explain the tradeoff between precision and recall.
8. Define F1-measure. What is the advantage of using F1-measure as an evaluation metric?

Lecture-09

1. Compare GDA and QDA in terms of parameter complexity.
2. Explain the naive Bayes assumption.
3. Gaussian Naive Bayes has linear decision boundary. True or False?
4. What is Laplace smoothing? Why do we need to smooth our Naive Bayes estimates?
5. Laplace smoothing is a biased smoothing. Justify.
6. What are the advantages of discriminative approach over generative approach for classification?
7. Explain the relationship between maximum likelihood and least squares.