

ISE – 364 / 464 Final Interview Question Bank

- 1) What is the difference between Lasso and Ridge Regression? What type of regularization does each use? How does lasso regression induce sparsity in its solution?
- 2) Explain what the bias-variance trade-off is. How do you handle a model that is underfitting. How do you handle a model that is overfitting?
- 3) What is the difference between a nonconvex, convex, and strongly convex problem in terms of what you can claim about the nature of the optimal solutions.
- 4) Explain how the definiteness of the Hessian matrix of a function has impacts on the optimality of a function. Explain how the gradient has impacts on the optimality of a function.
- 5) What theoretical assumption does linear regression make about the nature of the target variable distribution?
- 6) What are the main differences between gradient descent and Newton's method? How are they both derived? How are they connected?
- 7) Explain the difference between supervised learning and unsupervised learning.
- 8) Explain the difference between discriminative and generative models.
- 9) Explain the difference between the Expected Risk and the Empirical Risk. Why do we typically solve the empirical risk as opposed to the expected risk?
- 10) What are the Normal Equations? Give me an example of when using them could be problematic.
- 11) What can you tell me about the nature of the solution of a linear regression model (in terms of the Hessian of the loss function)?
- 12) What can you tell me about the nature of the solution of a logistic regression model (in terms of the Hessian of the loss function)?
- 13) What type of decision boundary is logistic regression learning?
- 14) What is the difference between linear and polynomial regression?
- 15) Explain what K-fold cross-validation is and how it can be used to generate confidence intervals on the expected performance of a model.
- 16) Why would one need to standardize data? Why would one need to normalize data? What types of models will require numeric features to be standardized and normalized?
- 17) Explain what Simpson's Paradox is.
- 18) Is Logicism a valid belief?
- 19) What quantities define a general hyperplane? What do you need to define a unique hyperplane?
- 20) What are two machine learning models that minimize the projected distance of datapoints to a hyperplane?
- 21) What are the main differences between a probability mass function and a probability density function?
- 22) Given two random variables  $X$  and  $Y$ , what does it mean if their covariance is non-zero? What does it mean geometrically?
- 23) Explain what the likelihood function is and what it is computing.
- 24) What does the softmax function do? What does it do to the input space?
- 25) Explain how the KNN machine learning algorithm works. If I want to use KNN to predict the target value of a new datapoint, how many distances do I need to compute? What effect does changing the size of  $K$  have on the decision boundary?
- 26) When training a classification tree, what are you trying to minimize? When training a regression tree, what are you trying to minimize?
- 27) What are the two main types of impurity metrics for a classification tree? Which is typically preferred in practice and why?

- 28) Explain an overview of the process used to train a decision tree.
- 29) Explain how a node chooses what feature to split on in a decision tree.
- 30) What are some ways to prevent decision trees from overfitting?
- 31) What is one reason why you may not want to use decision rules based on general hyperplanes in decision trees?
- 32) Why are multi-way decision splits not typically used in decision trees?
- 33) Explain what the purpose of bagging is. How does it work?
- 34) Explain what the purpose of boosting is. How does it work?
- 35) Explain what ensemble models are.
- 36) What is Stacking? How does it work?
- 37) Explain the Random Forest Algorithm.
- 38) Explain the general ideas behind the Gradient Boosting Algorithm.
- 39) What is the difference between Hard-Margin and Soft-Margin SVM?
- 40) Explain what the functional and geometric margins are. What is the difference between them?
- 41) What is a support vector in the context of SVMs?
- 42) Given an optimization problem, explain what the Lagrangian function is.
- 43) What is the difference between the primal and dual problem?
- 44) What are two reasons why we utilize the dual formulations of Hard and Soft SVM instead of their primal forms?
- 45) How can one learn nonlinear decision boundaries using a hyperplane?
- 46) What is the kernel trick (for SVM)?
- 47) Explain what forward and backward propagation are and how they are used to train a neural network.
- 48) Explain what the perceptron is. How is it connected to linear and logistic regression? How is it used to learn nonlinear patterns in MLPs?
- 49) Explain to me, step-by-step, how a neural network is trained via stochastic gradient descent.
- 50) How is the geometric margin formulation of SVM superior to the functional margin formulation?
- 51) What is the first-order necessary optimality condition? What is the second-order sufficient optimality condition? What are the KKT conditions?
- 52) Explain what Taylor Series are and how they are used in optimization.
- 53) Explain what Eigenvalues and Eigenvectors are. How are they interpreted geometrically?
- 54) In terms of formal systems, explain to me what the difference is between soundness, consistency, and completeness.
- 55) What is the shape of the 1-norm? What is the shape of the 2-norm? What is the shape of the infinity-norm?
- 56) Geometrically, what does the solution to a system of linear equations (given by  $Ax=b$ , where  $A$  is a matrix and  $x$  and  $b$  are vectors) represent? What about when the equality is replaced with an inequality?
- 57) In logistic regression, the target variable is modeled as a Bernoulli random variable. How is the target variable modeled in general multi-class logistic regression? What is the mathematical equation that is used as the predictive model for multi-class logistic regression?
- 58) Can we derive a closed form solution for linear regression? Can we derive a closed form solution for logistic regression?
- 59) Suppose I have a neural network. What is a loss function that I can use for a regression problem? What is a loss function that I can use for a binary classification problem? What is a loss function that I can use for a multi-class classification problem? Explain what the activation function is going to be in the final layer  $L$  for each of these scenarios.

- 60) What is a problem that both the sigmoid and hyperbolic tangent activation functions have in common when they are used in the hidden layers of a neural network? Does ReLU have a similar issue? What is the purpose of LeakyReLU?
- 61) Explain how the K-Means clustering algorithm works in detail.
- 62) What is the difference between K-means clustering and Gaussian mixture models? When are they equivalent?
- 63) Explain what PCA is and what it is doing.
- 64) What is the use of Dimensionality Reduction techniques?
- 65) What is the difference between label encoding and one-hot encoding. Why do we use them?