Machine Learning Question Bank

**Linear Regression, Logistic Regression, Gradient Descent and Regularization**

1. If the input data is of very high degree, how does regularization help?
2. Differentiate between Standard and Stochastic Gradient Descent.
3. Explain Gradient Descent algorithm for predicting parameters of Multivariate Linear Regression.
4. What do you understand by over-fitting of a classifier? How regularization can be used to tackle the problem of over-fitting?
5. Mention two disadvantages of the normal equation’s method, highlighting the instances where it cannot be used to determine the parameters of linear regression.
6. Differentiate between Standard Gradient Descent and Stochastic Gradient Descent.
7. Differentiate between Linear Regression and Logistic Regression.
8. If the input data is of very high degree, how does regularization help? Differentiate between scenarios where LASSO and ridge regression give different results.
9. Mention two disadvantages of the normal equations method, highlighting the instances where it cannot be used to determine the parameters of linear regression.
10. Mention the limitations of the gradient descent method in calculating the linear regression coefficients.
11. Describe three evaluation measures of regression models along with their formulas.
12. Given below is the data of five students who took a proficiency test as well as language course.

|  |  |  |
| --- | --- | --- |
| S.No. | X2 | Y (Output) |
| 1 | 95 | 85 |
| 2 | 85 | 95 |
| 3 | 80 | 70 |
| 4 | 70 | 65 |
| 5 | 60 | 70 |

1. Use the least square approximation to estimate the linear equation that best predicts language course performance, based on proficiency test scores?
2. Compute the sum of squared error (SSE) using the estimated model.
3. If a student scored 80 on the proficiency test, what marks would we expect her to obtain in the language course?
4. For values of b0 and b1 as -2.16 and 0.425 for categorical predictor variable X and a categorical response variable Y respectively, apply logistic regression to find P(Y=yes|X =yes) and P(Y=yes|X=no).
5. Find the linear regression coefficients using gradient descent method for the following dataset when learning rate = 0.2. Carry out the process for 2 iterations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 |
| Y | 3 | 4 | 5 | 4 | 6 |

1. Consider the dataset given below, which categorizes an article either as Technical (Class 1) or Non-Technical (Class 0) based on the time spent in reading (in Hours) and the number of sentences (in multiples of 1000) in that article.

|  |  |  |
| --- | --- | --- |
| Time (Hours) | Sentences (in multiple of 1000) | Y (Output) |
| 2.7 | 2.5 | 0 |
| 3 | 3 | 0 |
| 5.9 | 2.2 | 1 |
| 7.7 | 3.5 | 1 |

1. Using the above data, build a logistic regression model to predict the class of an article using gradient descent method. Assume 𝑙𝑒𝑎𝑟𝑛𝑖𝑛𝑔 𝑟𝑎𝑡𝑒=0.3. Further, in the first iteration the value of the coefficients is 0, and the *bias* is 1. Use two iterations of the gradient descent process to learn the model parameters.
2. Compute the error in prediction.
3. Use the above model to predict the article type of an article which requires **6.2 hours** of reading time and contains **3100 sentences**.
4. Explain Vanishing Gradient and Exploding Gradient problem.
5. Given the set of values X = (3, 9, 11, 5, 2)T and Y = (1, 8, 11, 4, 3)T. Evaluate the regression coefficients using ordinary least square method.
6. What is regularization? What is effect of the following on the model?
7. The regularization parameter (λ) is zero.
8. The regularization parameter (λ) is very large.
9. Find the least square regression line for the given dataset using normal equation method. Show computation at each step.

|  |  |  |
| --- | --- | --- |
| X1 | X2 | Y (Output) |
| 1 | 9 | 14 |
| 2 | 1 | 7 |
| 3 | 2 | 12 |
| 4 | 3 | 16 |
| 5 | 4 | 20 |

1. Consider the dataset given below having two input variables x1, x2 and one output variable y. Update the coefficients b0, b1 and b2 using gradient descent for the Logistic Regression model. Assume the learning rate = 0.3 and initial values of coefficients as b0 = -0.5, b1= 1 and b2= -1. Perform one iteration of gradient descent.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x1 | 1 | 3 | 8 | 6 |
| x2 | 2 | 5 | 3 | 2 |
| y | 0 | 0 | 1 | 1 |

1. Why cannot linear regression be used for classification? Explain with the help of an example.
2. What do you mean by polynomial regression? Explain it with an example.
3. The sales of a company (in million rupees) for each year are shown in the table below:

|  |  |
| --- | --- |
| x (year) | y (sales) |
| 2005 | 12 |
| 2006 | 19 |
| 2007 | 29 |
| 2008 | 37 |
| 2009 | 45 |
| 2010 | 49 |

1. Find the least square regression line y = ax+b
2. Use the least squares regression line as a model to estimate the sales of the company in 2013.
3. What is over-fitting in logistic regression? How can this problem be resolved?
4. Find the linear regression coefficients using gradient descent method for the following dataset when learning rate = 0.1. Carry out the process for 2 iterations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 |
| Y | 2 | 3 | 5 | 4 | 6 |

1. Explain how can logistic regression be used for solving more than two class problems?
2. What is the cost function for linear regression? Derive least square estimation of the coefficients?
3. Explain gradient descent method for obtaining the parameters of Logistic Regression.
4. Distinguish between overfitting and underfitting. How it can affect model generalization?
5. A classification model performs with a high accuracy on training data but generalizes poorly to new instances. Identify the problem and illustrate it with the help of a suitable figure. Enumerate three possible solutions to this problem.

**Naïve Bayes, Decision Tree and kNN**

1. Consider the following 10 training instances.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Color | Type | Mileage | Tested |
| 1 | Blue | Sports | Average | Yes |
| 2 | Blue | SUV | Average | No |
| 3 | Blue | Sports | High | Yes |
| 4 | Pink | Sports | High | No |
| 5 | Pink | Sports | Average | Yes |
| 6 | Pink | SUV | Average | No |
| 7 | Blue | SUV | High | Yes |
| 8 | Red | SUV | High | No |
| 9 | Red | SUV | Average | Yes |
| 10 | Red | Sports | Average | Yes |

Compute the following probabilities.

a. P(Blue|Yes) b. P(Sports|No) c. P(Red|Yes) d. P(Pink|Yes)

1. Mention two practical difficulties in applying the Bayesian methods.
2. (a) In case of a categorical attribute, if a particular conditional probability is zero, what strategy may be used to classify the data instance using naïve Bayes method.

(b) In case of a numerical attribute, how is the conditional probability calculated. Give an appropriate formulation.

1. Use Naïve Bayes classification rule for the following training data, predict whether an old student having high income and excellent credit rating will buy a computer or not.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Id | Age | Income | Student | Credit Rating | Buys Computer |
| 1 | Young | High | No | Fair | No |
| 2 | Young | High | No | Excellent | No |
| 3 | Middle | High | No | Fair | Yes |
| 4 | Old | Medium | No | Fair | Yes |
| 5 | Old | Low | Yes | Fair | Yes |
| 6 | Old | Low | Yes | Excellent | No |
| 7 | Middle | Low | Yes | Excellent | Yes |
| 8 | Young | Medium | No | Fair | No |
| 9 | Young | Low | Yes | Fair | Yes |
| 10 | Old | Medium | Yes | Fair | Yes |
| 11 | Young | Medium | Yes | Excellent | Yes |
| 12 | Middle | Medium | No | Excellent | Yes |
| 13 | Middle | High | Yes | Fair | Yes |
| 14 | Old | Medium | No | Excellent | No |

1. Identify the first splitting attribute for decision tree with the following dataset using ID3 algorithm:

|  |  |  |
| --- | --- | --- |
| **Field** | **Experience** | **Hired** |
| IT | Coding | No |
| IT | Coding | No |
| IT | Administration | Yes |
| IT | Administration | Yes |
| Business | Coding | Yes |
| Business | Coding | Yes |
| Business | Administration | No |
| Business | Administration | No |

1. Suppose that the probability of five events are P(first) = 0.5, P(second) = P(third) = P(fourth) = P(fifth) = 0.125. Calculate its entropy.
2. Consider the training data in the following table where Play is a class attribute.

|  |  |  |  |
| --- | --- | --- | --- |
| **Humidity** | **Sunny** | **Windy** | **Play** |
| L | N | S | N |
| H | N | W | Y |
| H | N | W | N |
| H | Y | S | Y |
| H | N | W | Y |
| L | Y | S | N |

Build a Naïve Bayes Classifier using the above data. Estimate the class Label for day where (Humidity = L, Sunny = N, Wind = W) using the above classifier.

**Neural Networks**

1. Draw a diagram for multilayer perceptron.
2. How can neural networks be used to tackle three class problems?
3. Show that single layer perceptron cannot solve XOR problem.
4. Give an expression of bipolar sigmoidal activation function. Also obtain the first derivative of the function.
5. How does an Artificial Neural Network resemble the functioning of brain?
6. Explain Backpropagation algorithm for Multilayer perceptron.
7. Train a neural network for the following data with X1 and X2 as inputs and Y as output.

|  |  |  |
| --- | --- | --- |
| X1 | X2 | Y (Output) |
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

1. Explain two methods of updating weights for a single layer perceptron.
2. Write the truth table of OR and solve it using single layer perceptron.
3. Give the neural network for implementing the NOR gate.
4. Design a neural network of one variable for the boolean function NOT. Show that the function is responding correctly with appropriate weight and bias.
5. Calculate the output y of a three inputs neuron with bias. The input feature vector is (𝑥1,𝑥2,𝑥3) = (0.8,0.6,0.4) and weight values are [𝑤1,𝑤2,𝑤3,𝑏]=[0.2,0.1,−0.3,0.35]. Use binary Sigmoid function as activation function.
6. Consider the following neural network with initial weights, biases and training input/outputs as mentioned.

0.05 0.15 0.40 0.01

I1 h1 o1

0.20 0.45

0.25 0.50 0.99

0.100.30h2 0.55 o2

I2

b1 b2

Given the inputs i1 = 0.05, i2 = 0.10, determine the values of output nodes o1 and o2. Also calculate the prediction error Etotal if the actual output values o1 and o2 are 0.01 and 0.99 respectively. Use sigmoid as the activation function for the hidden as well as the output layers.

**Bias-Variance**

1. Explain with diagrams
2. Fitting a simple model with high bias
3. Fitting a complex polynomial regression model with high variance
4. Near optimal polynomial model
5. Diagrammatically, show the a) high bias-low variance and b) low bias-high variance fit to a non-linear input data.

**Remaining**

1. List and explain applications of Machine learning.
2. Write Best Subset Selection algorithm. Comment on the complexity of this algorithm.
3. How is the cross validation technique useful while making the decision model? Why do we need to repeat the cross validations?
4. When is subset selection beneficial. State with a relevant example.
5. What do you mean by reinforcement learning? Given an example.
6. Define Concept Learning. How concept learning can be viewed as the task of searching?
7. Using an example discuss how new features can be constructed by forming Cartesian production of existing features. What are the implications of this approach?
8. Use K-means clustering to cluster the following data into two groups:

{2, 4, 10, 12, 3, 20, 30, 11, 25}

Assume cluster centroid are m1 = 2 and m2 =4. The distance function used is Euclidean distance.

1. Discuss the following with the help of examples – feature scaling, feature selection and feature extraction.
2. What is the difference between K-means clustering and K-Nearest Neighbor classifier?
3. Discuss the steps used by Principal Component Analysis to extract important features.
4. Assume a total of 1000 patients are tested for influenza; 900 are found to be healthy, while 100 are found to be sick. A test resulted in 60 being positive and 40 being negative for the sick persons. The same test was positive for 120 and negative for 780 in healthy adults. Construct a confusion matrix for the data and determine precision and recall.
5. State the mathematical formulation of the SVM problem. Give an outline of the method used for solving the classification problem using SVM.