WEEK-1 REPORT

DASARI SRINITH (CS21BTECH11015)

LINEAR REGRESSION:

1. What is the hypothesis function and the cost function in linear regression?

   Solution:
   The hypothesis function is ,

   $$h_\theta(x) = \theta^T x \tag{1}$$

   The cost function taken is ,

   $$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{2}$$

   where m is the number of training examples.

2. What is the LMS update rule or algorithm?

   Solution:
   The LMS algorithm is , to update all the parameters for j=1,2..n as,

   $$\theta_j := \theta_j - \alpha\frac{d}{d\theta_j}J(\theta) \tag{3}$$

   $$\theta_j := \theta_j - \alpha(h_\theta(x) - y)(x_j) \tag{4}$$

   This is for a single training example, for m training examples , we have to **repeat the following till convergence**

   $$\theta_j := \theta_j - \alpha\sum_{i=1}^{m}(h_\theta(x^{(i)} - y^{(i)})(x_j^{(i)}) \tag{5}$$

3. While applying batch gradient descent , is there a possibility that the minima is a local minima?

   Solution:
   No , the minima which we get on doing batch grade descent will be the global minima , as the cost function which we took is a convex quadratic function , so it converges to the global minimum.

4. What is the normal equations algorithm used as an alternative for batch grade descent?

   Solution: The normal equations uses the concept of matrices , using which we get the parameters of θ for which the cost function is minimum in a single shot, i,e, we have a closed form for the parameters.

$$\theta = (X^T X)^{-1} X^T \bar{y} \tag{6}$$

5. In the understanding of why least squares is considered for linear regression , what is the one word reason for why the errors are taken gaussian?

   Solution: Central limit theorem

LOGISTIC REGRESSION:

1. What is the hypothesis function and the cost function in logistic regression?

   Solution:
   The hypothesis function is ,

$$h_\theta(x) = \frac{1}{1 + \theta^T x} \tag{7}$$

   The cost function taken is ,

$$l(\theta) = \sum_{i=1}^{m} y^{(i)} \log(h(x^{(i)}) + (1 - y^{(i)})(\log(1 - h(x^{(i)})) \tag{8}$$

   where m is the number of training examples.

2. How are the updates for parameters done in logistic regression , how is it different with that of linear regression ?

   Solution: The parameters , for j=1,2,3...n , are updated as ,

$$\theta_j := \theta_j - \alpha(h_\theta(x) - y)(x_j) \tag{9}$$

   The update rule is similar to that of linear regression , the change is that , the $h_\theta(x)$ here is ,

$$h_\theta(x) = \frac{1}{1 + \theta^T x} \tag{10}$$

3. Is the decision boundary linear or nonlinear in logistic regression?

   Solution: The decision boundary is a Straight line i.e linear for a logistic regression

4. Write the generalized form for Newton's method in for logistic regression.

   Solution: The generalization of Newton's method to this multidimensional setting (also called the Newton-Raphson method) is given by

$$\theta := \theta - H^{-1}\nabla_\theta l(\theta) \tag{11}$$

   where $\nabla_\theta l(\theta)$, is, as usual, the vector of partial derivatives of $l(\theta)$ with respect to the θ_i's; and H is an n-by-n matrix (actually, n + 1-by-n + 1, assuming that we include the intercept term) called the Hessian, whose entries are given by

$$H_{ij} = \frac{d^2 l(\theta)}{d\theta_i d\theta_j} \tag{12}$$

5. Why isnt least squared error taken for logistic regression ?

   Solution: On considering the least squared error , we dont get a single global maximum , there will be local maximums due to which we dont get the accurate global maximum .

DECISION TREES

1. What are decision trees?

   Solution: Decision trees is a tool that uses a tree-like model of decisions and their possible consequences. If an algorithm only contains conditional control statements, decision trees can model that algorithm really well.Decision trees are used for classification and regression tasks.

2. How is cross-entropy loss defined ?

   Solution: Cross-entropy loss is defined as,

$$L_{cross} = \sum_c \hat{p_c} \log(\hat{p_c}) \tag{13}$$

3. How do you decide the stopping criteria for decision trees?

   Solution: A stopping criteria is needed when a decision tree learner runs, otherwise, it would overfit the data making it unable to give accurate predictions for new data. We could consider any of the following heuristics for stopping .

   (a) min leaf size
   (b) max depth
   (c) max number of nodes
   (d) min decrease in loss

(e) Pruning

4. What is Gini Index?

   Solution: It calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. It is given by ,

$$GI = 1 - \sum_{i=1}^{n} (p_i)^2 \tag{14}$$

   where p_i denotes the probability of an element being classified for a distinct class.

5. What is Greedy splitting?

   Solution: Greedy Splitting is also called Recursive Binary Splitting.In this procedure all the features are considered and different split points are tried and tested using a cost function. The split with the best cost (or lowest cost) is selected. All input variables and all possible split points are evaluated and chosen in a greedy manner (choosing the lowest value of cost possible).