

# Review Questions 1

Group 1

Chuan Su

Diego Alonso Guillen Rosaperez

November 10, 2019

1. All of these three statements are true about *Normal Equation*.

(a) We don't have to choose the learning rate.

(b) It becomes slow when number of features is very large.

(c) No need to iterate.

2.  $SSE = (-0.2)^2 + (0.4)^2 + (-0.8)^2 + (1.3)^2 + (-0.7)^2 = 3.02$

3. a and d are correct.

(a) In case of fewer observations, it is easy to overfit the data.

(d) In case of more observations, it is hard to overfit the data.

4. Two coefficients -  $w_0$  and  $w_1$  for  $y = w_1 * x + w_0$

5. *Cross-validation* is a technique useful for ensuring the effectiveness of a model, specially by mitigating overfitting. The training set is split into complementary training and validation subsets. Each model is trained against a different combination of these subsets and validated against the remaining parts. Once the model type and hyperparameters have been selected, a final model is trained using these hyperparameters on the full training set, and the test error is measured on the test set.

6. *Sigmoid*:

$$p1 = \text{sigmoid}(\theta h) = \frac{1}{1 + e^{-\theta h}}$$

$$p2 = 1 - \text{sigmoid}(\theta h) = \frac{1}{1 + e^{\theta h}}$$

*Softmax*  $k = 2$ :

$$p1 = \frac{e^{\theta_1 h}}{e^{\theta_1 h} + e^{\theta_2 h}} = \frac{1}{1 + e^{(\theta_2 - \theta_1)h}}$$

$$p2 = \frac{e^{\theta_2 h}}{e^{\theta_1 h} + e^{\theta_2 h}} = \frac{1}{1 + e^{(\theta_1 - \theta_2)h}}$$

Given  $\theta = \theta_1 - \theta_2$ , *softmax* function is equivalent to the *sigmoid* function with two classes.

7. **-log** ensures convexity of logistic regression cost function and convergence to a global minimum. Moreover, with **-log** the cost function output closes to 0 if the predicted value  $y$  is close to true values and large if the predicted value  $y$  is far from the true value.
8. (a) Cross-entropy and negative log-likelihood are the probabilistic interpretation of the logistic regression cost function.  
 (b) The equation of negative log-likelihood is the same as the logistic regression cost function. Therefore minimize the negative log-likelihood is to minimize the logistic regression cost function.  
 (c) Negative log-likelihood is also called cross-entropy, which is the logistic cost function. Cross-entropy quantify the difference (error) between the predicted distribution and the true distribution. Minimize the cross-entropy is to minimize the logistic regression cost function.
9. The receiver operating characteristic (ROC) curves summarize the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) for a model at different probability thresholds. Lowering the threshold classifies more items as positive, thus increasing both True Positive and False Positive. The higher the TPR is, the more FPR the classifier produces. To compute the points in an ROC curve, we need to evaluate the classification model, i.e. a logistic regression, many times with different threshold values.

$$TPR = \frac{TP}{TP+FN} \text{ and } FPR = \frac{FP}{FP+FN}$$

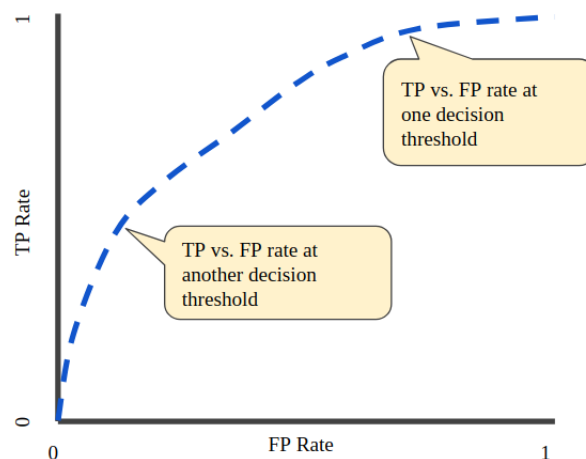


Figure 1: ROC curve from Google Developers

Google Developers: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>