National University of Singapore
School of Computing
CS3244: Machine Learning
Tutorial 3

Linear Classifiers and Logistic Regression

Colab Notebook: Linear Classifiers and Logistic Regression

1. **Logistic Regression**.

   Which of the following evaluation metrics cannot be applied in case of logistic regression output to compare with target? Explain your answer.

   (a) Accuracy
   (b) AUC-ROC
   (c) Log loss
   (d) Mean-Squared-Error

2. **Linear Regression Model Fitting**.

   You are given several data points as followed. Find a linear regression model that fits the data points best in terms of goodness-of-fit.

x_1	x_2	x_3	y
6	4	11	20
8	5	15	30
12	9	25	50
2	1	3	7

3. **Gradient of the Logistic Regression Cost Function**

   a) Derivative of the sigmoid function. The sigmoid function is given below. Express the derivative of the sigmoid function with respect to z in terms of the sigmoid function.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

   b) Derivative of the sigmoid function continued - What is the derivative of the function $log(1 - \sigma(z))$ with respect to z.

   c) Derivative of the cost function Given data points $\{(x_1, y_1), (x_2, y_2)....(x_m, y_m)\}$, the cost function for logistic regression is given by

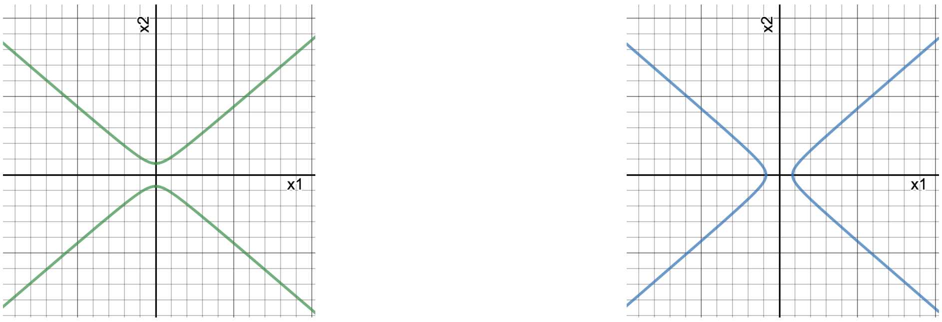
$$L(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y_i \ log(h_\theta(x_i)) + (1 - y_i) \ log(1 - h_\theta(x_i))$$

   where $h_\theta(x) = \sigma(\theta^T x)$. Derive the gradient of the cost function with respect to the parameters θ.

4. **Nonlinear Transformations.**

   Consider the feature transform $\Phi(1, x_1, x_2) = (1, x_1^2, x_2^2)$. Draw and show the boundary (not strictly) in $\mathcal{X}$ that a hyperplane $\hat{\theta}$ in $\mathcal{Z}$ correspond to under the following cases:

   (a) $\hat{\theta}_1 > 0, \hat{\theta}_2 < 0$.
   When $\hat{\theta}_0 > 0$, the plot shows a "North-South opening hyperbola"; when $\hat{\theta}_0 < 0$, it's a "East-West opening hyperbola" as shown in Figure 1.



(a) Borders when $\hat{\theta}_0 > 0$ (b) Borders when $\hat{\theta}_0 < 0$

Figure 1: Borders for the two cases of $\hat{\theta}_0$

   (b) $\hat{\theta}_1 > 0, \hat{\theta}_2 = 0$.

   (c) $\hat{\theta}_1 > 0, \hat{\theta}_2 > 0, \hat{\theta}_0 < 0$.

   (d) $\hat{\theta}_1 > 0, \hat{\theta}_2 > 0, \hat{\theta}_0 > 0$.

5. [**]**The Hat Matrix.**

   The hat matrix is an integral part of understanding linear regression. We see the hat matrix makes an appearance in the analytical, closed form solution of linear regression, as it contains the pseudo inverse $(\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\top$. *Let's look at its properties.*

   Consider the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\top$, where $\mathbf{X}$ is an m by $n+1$ matrix, and $\mathbf{X}^\top\mathbf{X}$ is invertible.

   (a) Show that $\mathbf{H}$ is symmetric.

   (b) Show that $\mathbf{H}^k = \mathbf{H}$, for any positive integer k.

   (c) If $\mathbf{I}$ is the identity matrix of size N, show that $(\mathbf{I} - \mathbf{H})^K = \mathbf{I} - \mathbf{H}$ for any positive integer K.

   (d) Show that $trace(\mathbf{H}) = n+1$, where the *trace* (*tr*) is the sum of diagonal elements. [*Hint:* $trace(\mathbf{AB}) = trace(\mathbf{BA})$.]