50.021 Artificial Intelligence

Theory Homework 2

Due: every Tuesday, 6PM

[Q1]. Find *any* separating hyperplane equation for these three sample points: $\boldsymbol{x}_1 = (-2, 2), y_1 = -1$, $\boldsymbol{x}_2 = (4, 0), y_2 = -1$, and $\boldsymbol{x}_3 = (-2, -3), y_3 = +1$. Draw (by hand) or plot (using Python, see matplotlib) the result.

[Q2]. Find by hand the value of optimum weights $\hat{\boldsymbol{w}}$ and bias $\hat{b}$ using linear regression for the four following sample points: $\boldsymbol{x}_1 = (1, 0), y_1 = +1$, $\boldsymbol{x}_2 = (2, 3), y_2 = +1$, $\boldsymbol{x}_3 = (3, 4), y_3 = -1$, $\boldsymbol{x}_4 = (3, 2), y_4 = -1$. Show your working. If you have to do matrix inversion, then you can use a program to compute the matrix inverse. Answer the following question: How can you obtain a solution via the closed-form method (that is not by using gradients) which does not use an inverse matrix $(X^T X)^{-1}$ to obtain the weight vectoe w?

[Q3] In logistic regression, we model the probability of a label $y \in \{-1, 1\}$ given a sample point $\boldsymbol{x} \in \mathbb{R}^2$ as,

$$P(Y = y|\boldsymbol{x}) = \frac{1}{1 + e^{-y(\boldsymbol{w}\cdot\boldsymbol{x}+c)}},$$

where c is the bias term. Answer the following questions,

1. Obviously the function above alone is not linear. However, given that after we know what the values $\boldsymbol{w}$ are (after training), *prove* that the *decision boundary* obtained from logistic regression is linear, i.e.: the decision boundary is a hyperplane with equation $\boldsymbol{w} \cdot \boldsymbol{x} = 0$. *Hint: Since $\boldsymbol{x}$ is two-dimensional, the decision boundary is the set of $\boldsymbol{x}$ such that $P(Y = -1|\boldsymbol{x}) = P(Y = 1|\boldsymbol{x})$.*

2. Suppose we have obtained the optimum weights, $\hat{w} = (w_1, w_2)$ and bias $\hat{c}$ in the case of 2-dimensional data points. The decision boundary is supposedly a line that separate points with positive and negative labels, which has an equation in the form of $x_2 = mx_1 + k$. Obtain an expression for m and k in terms of w_1, w_2 and $\hat{c}$. When this expression of the form $x_2 = mx_1 + k$ is not defined?

[Q4]. The following is the regularized logistic regression loss function.

$$L(w) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log(h_w(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)}))\right] + \frac{\lambda}{2}||w||^2$$

where $h_w(x) = \frac{1}{1+e^{-w\cdot x}}$. Compute the gradient of the loss with respect to w and write down the gradient descent update equation. Show all steps clearly.