Machine Learning Homework # 1

Arvi Gjoka

1. In creating machines for the perceptron, we define the new input vector as $\tilde{x} = [x; 1]$, where the new entry acts as a constant offset for our weighted fitting $w^T\tilde{x} = \sum_i^{n-1} w_i x_i + w_n$. We need to do this in case our data is offset from the origin that our fitting needs to include a constant term to translate the line fit around the data. This is similar to the constant term α_0 in a polynomial fit of data with the form $f(x) = \alpha_2 x^2 + \alpha_1 x + \alpha_0$. Consider two points on the x axis equidistant from one another. Drawing a line of best fit is trivial using no offset. However, if the two points are translated to the right more than half their distance, fitting a line to them becomes impossible without an offset from the axis. When you have many points clustered around some offset from the origin, the constant offset is vital.

2. The distance function has a trivial solution when $w^T\tilde{x} = 0$, which means that the augmented input vector and the weight vector are perpendicular. The boundary line is defined as $w^T\tilde{x} = \sum_i^{n-1} w_i x_i + w_n = 0$ and a trivial solution corresponds to a point occurring on this line. One solution to this is to randomize the initial weight vector and start the learning process again, with the hopes that as the weight vector changes again, the decision boundary does not sweep the points. Another thing would be to skip points that are on the boundary line and learn them at the end, when the contribution is not 0.

3.

$$\begin{gathered}
\text{First, we have} \\
M(x,w) = \frac{1}{1+e^{-w^T\tilde{x}}} \\
\frac{\partial M}{\partial w} = \tilde{x}\frac{1}{1+e^{-w^T\tilde{x}}}\frac{e^{-w^T\tilde{x}}}{1+e^{-w^T\tilde{x}}} \\
\frac{\partial M}{\partial w} = \tilde{x}M(1-M) \\
\text{Given this, we can find} \\
\frac{\partial D}{\partial w} = -\left(\frac{y}{M}\frac{\partial M}{\partial w} - \frac{1-y}{1-M}\frac{\partial M}{\partial w}\right) \\
\frac{\partial D}{\partial w} = -\tilde{x}(y - yM - M + yM) \\
\frac{\partial D}{\partial w} = \tilde{x}(M(x,w) - y)
\end{gathered}$$