Machine Learning: Homework #5

Due on November 27, 2017 at 09:59am

Professor Dr. Stephan Guennemann

Marc Meissner - 03691673

Problem 1

Show that if their convex hulls intersect, the two sets of points can not be linearly separable.

Solution
Using the augmented w and x vectors by integrating w_0 into w:

$$Z(x) = w^T x = w^T (\sum_{i=0}^{N} \alpha_n x_n) = \sum_{i=0}^{N} \alpha_n (w^T x_n)$$

Similarly, for y:

$$Z(y) = \sum_{m=0}^{M} \beta_m (w^T y_m)$$

Intersecting them at a point xy would lead to:

$$Z(xy) = \sum_{i=0}^{N} \alpha_n (w^T x_n) \stackrel{!}{=} \sum_{m=0}^{M} \beta_m (w^T y_m)$$

Since $\alpha_i, \beta_i > 0$ for all i and $w^T x_n > 0$ as well as $w^T y_n < 0$, this formula has no solution. One side is negative, while the other is positive. Hence, it is not possible for the two sets of points to be linearly separable, if the convex hulls intersect.

Problem 2

Show that for a linearly separable data set, the maximum likelihood solution for the logistic regression model is obtained by finding a vector w whose decision boundary $w^T x = 0$ separates the classes and then taking the magnitude of w to infinity. Assume that w contains the bias term. How can we prevent this?

Solution
Imagine two classes $y \in \{0, 1\}$ and the set of points x_1, x_2 for which $w^T x_1 > 0$ and $w^T x_2 < 0$ holds. Then:

$$\lim_{w\to\infty} \sigma(w^T x_1) = \lim_{w\to\infty} \frac{1}{1 + e^{-w^T x_1}} = 1 \text{ and}$$
$$\lim_{w\to\infty} \sigma(w^T x_2) = \lim_{w\to\infty} \frac{1}{1 + e^{-w^T x_2}} = 0.$$

This corresponds to the maximum likelihood solution as it hard thresholds the weighted points. By letting the weights w go to infinity, the sigmoid function effectively becomes a Heaviside/Ramp function.
This can be avoided by applying regularization to keep the weights small.

Problem 3

Which basis function $\Phi(x_1, x_2)$ makes the data in the example linearly separable (crosses in one class, circles in the other)?

Problem 3 continued on next page. . .

Solution
$\Phi(x_1, x_2) = x_1 x_2$

Problem 4

The decision boundary for a linear classifier on two-dimensional data crosses axis x_1 at 2 and x_2 at 5. Write down the general form of this linear classifier model with a bias term (how many parameters do you need, given the dimensions?) and calculate possible coefficients (parameters).

Solution
Since $x \in \mathbb{R}^2$, $w^T \in \mathbb{R}^2$ and $w_0 \in \mathbb{R}^1$. The linear model is:

$$y(x) = w^T x + w_0 = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + w_0$$

The following three points must satisfy the formula:

$$y\left[\begin{pmatrix} 0 \\ 5 \end{pmatrix}\right] = 5w_2 + w_0 \overset{!}{=} 0$$
$$y\left[\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right] = 2w_1 + w_0 \overset{!}{=} 0$$
$$y\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right] = w_0 \overset{!}{<} 0$$

It follows that $w_2 \overset{!}{=} \frac{2}{5} w_1$. We can choose e.g. $w_0 = -2$ arbitrarily. Then $w_1 = 1$ and $w_2 = \frac{2}{5}$.
Final results:

$$w^T = \begin{pmatrix} 1 & \frac{2}{5} \end{pmatrix}$$
$$w_0 = -2$$