Machine Learning and Applications

Assignment 4

1. (2) Given a set of data points $\{\mathbf{x}_n\}$, we can define the convex hull to be the set of all points $\mathbf{x}$ given by

$$\mathbf{x} = \sum_n \alpha_n \mathbf{x}_n$$

where $\alpha_n \geq 0$ and $\sum_n \alpha_n = 1$. Consider a second set of points $\{\mathbf{y}_n\}$ together with their corresponding convex hull. By definition, the two sets of points will be linearly separable if there exists a vector $\widehat{\mathbf{w}}$ and a scalar w_0 such that $\widehat{\mathbf{w}} \cdot \mathbf{x}_n + w_0 > 0$ for all $\mathbf{x}_n$, and $\widehat{\mathbf{w}} \cdot \mathbf{y}_n + w_0 < 0$ for all $\mathbf{y}_n$. Show that if their convex hulls intersect, the two sets of points cannot be linearly separable, and conversely that if they are linearly separable, their convex hulls do not intersect.

2. (2) Show that for a linearly separable data set, the maximum likelihood solution for the logistic regression model is obtained by finding a vector $\mathbf{w}$ whose decision boundary $\mathbf{w}^T\phi(\mathbf{x}) = 0$ separates the classes ($\phi(x)$ is transformation of the input vectors), and then taking the magnitude of $\mathbf{w}$ to infinity.

3. (2) Show that the Hessian matrix $\mathbf{H}$ for the logistic regression model, given by

$$\mathbf{H} = \nabla\nabla E(\mathbf{w}) = \sum_{n=1}^{N} y_n(1-y_n)\phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T = \mathbf{\Phi}^T\mathbf{R}\mathbf{\Phi},$$

is positive definite. Here $E(\mathbf{w})$ is an error function:

$$E(\mathbf{w}) = -\sum_{n=1}^{N}(t_n \ln y_n + (1-t_n)\ln(1-y_n))$$

t_n is the actual class label, y_n is the output of the logistic regression model for input vector $\mathbf{x}_n$:

$$y_n = \sigma(\mathbf{w}^T\phi(\mathbf{x}_n)),$$

$\mathbf{R}$ is a diagonal matrix with elements $y_n(1 - y_n)$ and $\mathbf{\Phi}$ is a matrix whose n-th row is given by $\phi(\mathbf{x}_n)^T$

Hence show that the error function is a concave function of $\mathbf{w}$ and that it has a unique minimum.