Machine Learning Assignment 5

Shivangi Aneja

27-November-2017

Problem 1

$\mathcal{X} = \{x_i\}_{i=1}^{N} \qquad co\mathcal{X} = \{x : x = \sum_i \alpha_i x_i, \alpha_i \geq 0, \sum_i \alpha_i = 1\}$

$\mathcal{Y} = \{y_j\}_{j=1}^{M} \qquad co\mathcal{Y} = \{y : y = \sum_j \beta_j y_j, \beta_j \geq 0, \sum_j \beta_j = 1\}$

Two set of points $co\mathcal{X}$ and $co\mathcal{Y}$ are linearly separable if there exists a vector $\mathbf{w}$ and a scalar w_0 such that
(1) $\mathbf{w}^T x_i + w_0 > 0, \qquad x_i \in \mathcal{X}$
(2) $\mathbf{w}^T y_j + w_0 < 0, \qquad y_j \in \mathcal{Y}$

Suppose we have $\mathbf{z}$ that lies in both the regions, so linear discriminant for $\mathbf{z}$ can be written as ,
$y(z) = \sum_i \alpha_i(\mathbf{w}^T x_i + w_0) = \sum_j \beta_j(\mathbf{w}^T y_j + w_0)$
So we have ,
(3) $\sum_i \alpha_i(\mathbf{w}^T x_i + w_0) > 0 \Rightarrow \sum_i \alpha_i \mathbf{w}^T x_i + \sum_i \alpha_i w_0 > 0 \Rightarrow \sum_i \alpha_i \mathbf{w}^T x_i > -w_0$
(4) $\sum_j \beta_j(\mathbf{w}^T y_j + w_0) < 0 \Rightarrow \sum_j \beta_j \mathbf{w}^T y_j + \sum_j \beta_j w_0 < 0 \Rightarrow \sum_j \beta_j \mathbf{w}^T y_j < -w_0$

But both (3) and (4) are contradicting, as $y(z)$ has to be greater and less than zero at the same time, which is not possible for one single data point.

Hence , the convex hulls do not intersect

Problem 2

The negative log likelihood for Logistic Regression is given as :
$E(w) = -\sum_{i=1}^{N} [y_i ln\ \sigma(\mathbf{w^T x_i}) + (1 - y_i) ln\ \sigma(1 - \mathbf{w^T x_i})]$
Taking Gradient, we obtain
$\nabla_w E(w) = \sum_{i=1}^{N} [\sigma(\mathbf{w^T x_i}) - y_i] x_i$

This solution is maximized when $\boxed{\sigma(\mathbf{w^T x_i}) = y_i}$ for all i (data points)
This occurs when sigmoid function is saturated, i.e $\sigma = 0.5$
Maximum Likelihood exhibits severe overfitting for linearly separable data sets.This is because ML solution occurs when hyperplane corresponding to $\sigma = 0.5$ separates the 2 classes.
$\sigma = 0.5 \Rightarrow \mathbf{w^T x} = 0 \Rightarrow w \rightarrow \infty$
In this case, the logistic sigmoid function becomes infinitely steep in feature space, corresponding to a Heaviside step function, so that every training point from each class k is

assigned a posterior probability $p(C_k|x) = 1$.

How To avoid : This can be avoided by including of a prior and finding a MAP solution for w, or by adding a regularization term to the error function.

Problem 3

The equation for the figure is given by : $x_1^2 + x_2^2 = 1$

For Blue Cross, we have ,
(1) $0 \leq x_1 \leq 1$ and $-1 \leq x_2 \leq 0$
(2) $-1 \leq x_1 \leq 0$ and $0 \leq x_2 \leq 1$
Thus , using (1) and (2), $x_1 x_2 \leq 0$ for this case

For Black Circle, we have ,
(3) $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 1$
(4) $-1 \leq x_1 \leq 0$ and $-1 \leq x_2 \leq 0$
Thus , using (3) and (4), $x_1 x_2 \geq 0$ for this case

$\boxed{\phi(x_1, x_2) = x_1 x_2}$

Thus function $\boxed{\phi(x_1, x_2) = x_1 x_2}$ makes the data points linearly separable with $\phi > 0$ for circles and $\phi < 0$ for Crosses

Problem 4

The decision boundary crosses x_1 at 2 and x_2 at 5.

The general form of the decision boundary for linear classifier is given as :
$\boxed{\mathbf{w^T x} + w_0 = 0}$

Here we have D = 2 , so
$\boxed{w_1 x_1 + w_2 x_2 + w_0 = 0}$

For $x_1 = 2$, we have,

$2w_1 + w_0 = 0 \Rightarrow \boxed{w_1 = \frac{-w_0}{2}}$

For $x_2 = 5$, we have,

$$5w_2 + w_0 = 0 \Rightarrow \boxed{w_2 = \frac{-\,w_0}{5}}$$