Machine Learning HW 9

Support Vector Machine

Bui Phuong Thao - 11215341 - DSEB 63

1 Exercise 1

Assume that the given data set is linearly seperable, a new data point could be classified according to

$y(x) = w^T\phi(x) + b$ where w = $\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$

$w^T x_i + b \geq 1; y_i = 1$
$w^T x_i + b \leq -1; y_i = -1$
We need to find w and b such that the above conditions are satisfied and the margin is maximized
The distance from a point (x_0, y_0) to a line $(Ax + By + c = 0)$ is:
$\frac{|Ax_0+By_0+C|}{\sqrt{A^2+B^2}} = \frac{|wx+b|}{\|w\|}$
The margin is given by:
$min_{x_i,y_i=-1} = \frac{|wx+b|}{\|w\|} + min_{x_i,y_i=1} = \frac{|wx+b|}{\|w\|} = \frac{2}{\|w\|}$
To maximize $\frac{2}{\|w\|}$ we need to minimize $\| w \|$
$argmin_{w,b}\frac{1}{2} \| w \|^2$
This is a constrained optimization problem:
minJ(w) = $argmin_{w,b}\frac{1}{2} \| w \|^2$
st $y_i(w^T\phi(x) + b) \geq 1)$, i = 1,2,3,...N
Use Larange multiplier method:
The primal problem:
L(w,b,a) = $\frac{1}{2} \| w \|^2 - \sum_{n=1}^{N} a_n y_n (w^T x_n + b) - 1$ where a = $[a_1, a_2, , , , , a_n]^T$
Take the derivative of L with respect to w and b:
$\frac{\partial L(w,b,a)}{\partial w} = w - \sum_{n=1}^{N} a_n y_n x_n \rightarrow \sum_{n=1}^{N} a_n y_n x_n = w$

$\frac{\partial L(w,b,a)}{\partial b} = \sum_{n=1}^{N} a_n y_n \rightarrow \sum_{n=1}^{N} a_n y_n = 0$
The previous primal problem could be transformed into the dual:
$minmax_{w,b,a\geq 0}(\frac{1}{2} \parallel w \parallel^2 - \sum_{n=1}^{N} a_n(y_n(w^T x_n + b)) - 1)$
Instead of minimizing the primal over w,b, we can maximize the dual over a
$L(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m y_n y_m x_n^T x_m$
st $\sum a_n y_n = 0$ and $a_n \geq 0$
$L(a) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m y_n y_m k(x_n, x_m)$
Define $k(x, x') = \phi(x^T)\phi(x')$
The kernel function is involved because it gives measure of similarity, it is positive definite so that L(a) is bounded below
L(a) must also satisfy the KKT conditions:

$$a_n \geq 0$$
$$y_n(w^T x_n + b) - 1 \geq 0$$
$$a_n(y_n(w^T x_n + b) - 1) = 0$$

If $a_n = 0 \rightarrow datapointhasnorolethepredictions$
$Ify_n(w^T x_n + b) = 1$ then the data point is on the maximum margin hyperplanes
Thus any support vector x_n satisfies: $y_n(w^T x_n + b) = 1$
$\rightarrow y_n(\sum_{m\in S} a_m y_m k(x_n, x_m) + b) = 1$
$\rightarrow b = 1\frac{}{N_S} . \sum_{n\in S}(t_n - \sum_{m\in S} a_m y_m k(x_n, x_m))$