

CS771A : Machine Learning: Tools, Techniques and Applications

Assignment 2

Divyanshu Shende
Roll No. 13264

March 18, 2016

1 Spam Filter using SVM

Five Fold Cross validation was used. Here are the results:

Accuracy = [0.98965517, 0.99309154, 0.99307958, 0.98788927, 0.98961938]

Precision = [1, 0.9893617, 1, 1, 1]

Recall = [0.93814433, 0.96875, 0.95833333, 0.92708333, 0.9375]

Average Accuracy = 0.990666988968

Average Precision = 0.997872340426

Average Recall = 0.945962199313

2 Connect Four

Here again Five Fold Cross Validation was used in both cases.

1. One vs Rest : The accuracies were as follows : 0.774422735346, 0.809058614565, 0.710923623446, 0.759029011249, 0.681027463173. The average accuracy was 0.74689228955.
2. One vs One : The accuracies were as follows : 0.681172291297, 0.673401420959, 0.556912374186, 0.642391947898, 0.556295802798. The average accuracy was 0.62203476742.

A3) (a) If k_1 is a kernel on \mathcal{P} , then $K(\vec{x}, \vec{z}) = e^{k_1(\vec{x}, \vec{z})}$ is a kernel, psd
 k_1 is a ~~kernel~~ $\Rightarrow K^{(1)} = k_1 \circ k_1$ is a psd (Henderson Prod.)
 (Schur's Thm.)
 $\Rightarrow K^{(2)} = k_1 \circ k_1$ is a kernel fn.

In general let $K^{(n)} = \underbrace{k_1 \circ k_1 \dots \circ k_1}_{n \text{ times}}$

$\Rightarrow K^{(n)}$, where, $K_{i,j}^{(n)} = (k_{1,i,j})^n$, is psd.

~~Let~~ $\&$

$K^{(n)}$ is p.s.d. $\forall n \geq 1$

$$\Rightarrow \sum_i \sum_j x_i x_j K_{i,j}^{(n)} \geq 0$$

$$= \sum_i \sum_j x_i x_j \frac{K_{i,j}^{(n)}}{n!} \geq 0$$

Now, summing over n gives,

~~Let~~

$$\sum_i \sum_j x_i x_j \sum_{n=1}^{\infty} \frac{K_{i,j}^{(n)}}{n!} \geq 0$$

\Rightarrow

$$\sum_i \sum_j x_i x_j K e^{k_1(\vec{x}, \vec{z})} \geq 0$$

$\therefore e^{k_1(\vec{x}, \vec{z})}$ is a kernel fn.

(2)

(b) Take ϕ to be

$$\phi(\vec{x}) = \frac{e^{\|\vec{x}\|^2}}{\|\vec{x}\|^2} \cdot \vec{x}$$

$$\langle \phi(\vec{x}), \phi(\vec{y}) \rangle = \left\langle \frac{e^{\|\vec{x}\|^2}}{\|\vec{x}\|^2} \cdot \vec{x}, \frac{e^{\|\vec{y}\|^2}}{\|\vec{y}\|^2} \cdot \vec{y} \right\rangle$$

$$= \frac{e^{\|\vec{x}\|^2 + \|\vec{y}\|^2}}{\|\vec{x}\|^2 \cdot \|\vec{y}\|^2} \vec{x}^T \vec{y}$$

$$= k(x, y)$$

(c)

A4) a) The problem is

$$\min_{w, \xi} L = \sum_{i=1}^n \xi_i^2 = \sum_{i=1}^n (y_i - w^T x_i)^2$$

subject to $\|w\|_2 \leq B$

The dual is,

$$\text{maximize}_\alpha L = \sum (y_i - w^T x_i)^2 + \alpha (\|w\|_2^2 - B^2)$$

where $\alpha \geq 0$.

Now, we apply KKT conditions!

$$\frac{\partial}{\partial w} (L) = 0$$

$$\Rightarrow \sum_{i=1}^n 2(y_i - w^T x_i) x_i = 2\alpha w$$

Solving, we get

$$\vec{w} = \alpha (X^T X + I)^{-1} X^T y$$

$$\text{where } X_{n \times n} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

$$\text{and } y_{n \times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

A few vals. of α may be forbidden so that $(X^T X + \alpha I)^{-1}$ exists. Also \vec{w} is a fun. of α

*(b). Yes. Indeed the problem has an equivalent of support vectors. Note that not all points affect the ~~extreme~~ final val of \bar{w} . Therefore, the ones that do, form the equivalent of support vectors.

(c) The dependence of \bar{w} on d is a problem. This did not occur in case of SVM's.