Machine Learning Mandatory Assignment

May 2022

1 Kernels and mapping functions

a. The formula of the Binomial theorem is: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$.

$\mathrm{K}(\vec{x}, \vec{y}) = (\vec{x}\vec{y}+1)^3 = ((x_1y_1 + x_2y_2)+1)^3 = (x_1y_1+x_2y_2)^3 + 3(x_1y_1+x_2y_2)^2 + 3(x_1y_1+x_2y_2)+1$.

$(x_1y_1 + x_2y_2)^3 = x_1^3y_1^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + x_2^3y_2^3$

$3(x_1y_1+x_2y_2)^2 = 3(x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2)$

$\mathrm{K}(\vec{x}, \vec{y}) = x_1^3y_1^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + x_2^3y_2^3 + 3(x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2) + 3(x_1y_1+x_2y_2) + 1 = x_1^3y_1^3 + x_2^3y_2^3 + 3x_1^2x_2y_1^2y_2 + 3x_1x_2^2y_1y_2^2 + 3x_1^2y_1^2 + 3x_2^2y_2^2 + 6x_1x_2y_1y_2 + 3x_1y_1 + 3x_2y_2 + 1$.

$$\Downarrow$$

$\psi(\vec{x}) = \psi(x_1, x_2) = (x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{6}x_1x_2, \sqrt{3}x_1, \sqrt{3}x_2, 1)$.

b. Full Rational Variety.

c. $\psi(\vec{x}) \cdot \psi(\vec{y})$ uses 46 multiplication operations. $\mathrm{K}(\vec{x}, \vec{y})$ uses 4 multiplication operations. Thus, the difference is 42.

2 Lagrange multipliers

The Lagrangian is : $L(x, y, \lambda) = 2x - y + \lambda(\frac{x^2}{2} + y^2 - 1)$

The derivatives are: $\frac{\partial L}{\partial \lambda} = \frac{x^2}{4} + y^2 - 1$

$\frac{\partial L}{\partial x} = 2 + \frac{\lambda x}{2}$

$\frac{\partial L}{\partial x} = -1 + 2\lambda y$

Next we will solve the following equations:

$\frac{x^2}{2} + y^2 - 1 = 0$

$2 + \frac{\lambda x}{2} = 0$

$-1 + 2\lambda y = 0$

The solution to this system of equations is equal to:
$\lambda = +\frac{\sqrt{17}}{2}$, $x = \frac{-8}{\sqrt{17}}, y = \frac{1}{\sqrt{17}}$.

$\lambda = -\frac{\sqrt{17}}{2}, x = \frac{8}{\sqrt{17}}, y = \frac{-1}{\sqrt{17}}$.

$f(\frac{-8}{\sqrt{17}}, \frac{1}{\sqrt{17}}) = -\sqrt{17}$, minimum value.

$f(\frac{8}{\sqrt{17}}, \frac{-1}{\sqrt{17}}) = \sqrt{17}$, maximum value.

3 PAC Learning

Given the three constraints in h(r), plug the vectors in and form a formula that describes a line:

$(x, y) \cdot u \leq r \implies (x, y) \cdot (\frac{\sqrt{3}}{2}, \frac{1}{2}) \leq r \implies \frac{\sqrt{3}}{2}x + \frac{y}{2} \leq r \implies \frac{y}{2} \leq \frac{\sqrt{3}}{2}x + r$

$(x, y) \cdot v \leq r \implies (x, y) \cdot (\frac{\sqrt{3}}{2}, \frac{-1}{2}) \leq r \implies \frac{\sqrt{3}}{2}x + \frac{-y}{2}y \leq r \implies \frac{y}{2} \geq \frac{\sqrt{3}}{2}x - r$

$(x, y) \cdot w \leq r \implies (x, y) \cdot (-1, 0) \leq r \implies x \leq r \implies x \geq -r$

With these three inequalities:

$\frac{y}{2} \leq \frac{\sqrt{3}}{2}x + r$

$\frac{y}{2} \geq \frac{\sqrt{3}}{2}x - r$

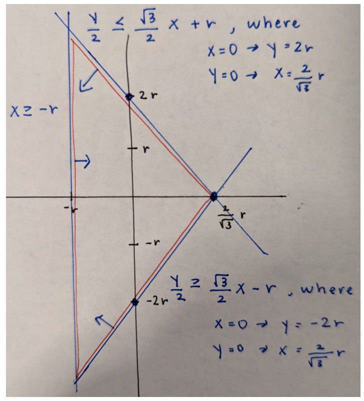
$x \geq -r$

They can be broken down into points when plugging in x=0 and y=0:

In $\frac{y}{2} \leq \frac{\sqrt{3}}{2}x + r$: x=0 yields $y = 2r$ and y=0 yields $x = \frac{2}{\sqrt{3}}r$

In $\frac{y}{2} \geq \frac{\sqrt{3}}{2}x - r$: x=0 yields $y = -2r$ and y=0 yields $x = \frac{2}{\sqrt{3}}r$

Plotting these points, we see an origin centered upright equilateral triangle:



Using this graph, we can define three regions so that each region B_i has probability $P(B_i) = \frac{\varepsilon}{3}$. Given a data set with. instances that visits each region, the error of this triangle defined by this data set will be less than ε.
A polynomial sample complexity algorithm L that learns C using H would be to take the maximum of of all m samples in $\{X_i, Y_i\}$ that satisfy the conditions in the inequalities. Since all m points are within that triangle, taking the max point that meets the conditions is the algorithm needed for L to learn C using H. The time complexity of this algorithm is O(m), since it must iterate through each of the m points in the data set.

The sample complexity will be found with the knowledge that $Error(h, c) < \varepsilon$. To not be ε_{bad}, the points need to have never visited at least one of the regions. The set of bad data points is contained in the union of all data points that don't visit region 1, region 2 and region 3.

$P(\{D \in X^m : Err(L(D), C) \geq \varepsilon\} \leq \sum_{i=1}^{4}(P(X - B_i))^m$

$\sum_{i=1}^{4}(P(X - B_i))^m \leq 3(1 - \frac{\varepsilon}{3})^m \leq 3e(\frac{-m\varepsilon}{3})$

Set this to be $\leq 1 - \delta$ to limit the bound and then solve for m:

$3e(\frac{-m\varepsilon}{3}) \leq 1 - \delta$

$e(\frac{-m\varepsilon}{3}) \leq 3(1 - \delta)$ and then $ln()$ the entire formula

$\frac{-m\varepsilon}{3} \leq 3ln(1 - \delta))$

$m \geq \frac{3}{\varepsilon} ln(\frac{3}{\delta})$ is the sample complexity

Therefore, it can be concluded that L is a PAC learnable algorithm.

4 Classification Error

$se = \sqrt{\frac{\hat{p}\times(1-\hat{p})}{n}} = \sqrt{\frac{0.2\times0.8}{1000}} = 0.0126$

$2 \cdot se = 2 \cdot 0.0126 = 0.025 = 2.5\%$

With 95% confidence, the generalization error interval is :

$(20\% - 2.5\%, 20\% + 2.5\%)$.

5 SVM

The Cs values are Cs = [0.0005, 0.001, 0.005, 0.01, 0.1, 1, 100].

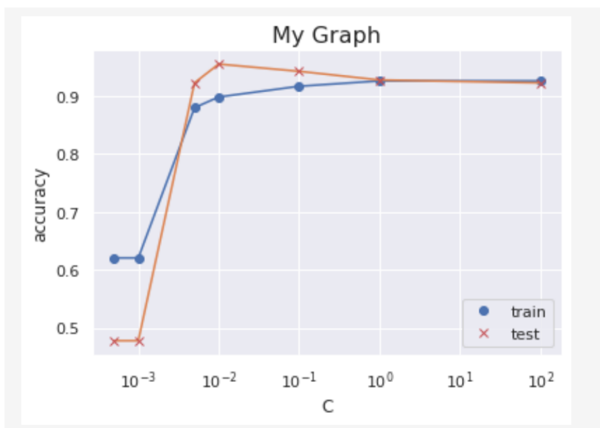


Figure 1: SVM My Graph