University of Illinois Spring 2018

CS 446: Machine Learning
Homework

Due on Tuesday, Feb 13, 2018, 11:59 a.m. Central Time

1. **[10 points]** SVM Basics
Consider the following dataset $\mathcal{D}$ in the two-dimensional space; $\mathbf{x}^{(i)} \in \mathbb{R}^2$ and $y^{(i)} \in \{1, -1\}$

i	$\mathbf{x}_1^{(i)}$	$\mathbf{x}_2^{(i)}$	$y^{(i)}$
1	-1	3	1
2	-2.5	-3	-1
3	2	-3	-1
4	4.7	5	1
5	4	3	1
6	-4.3	-4	-1

Recall a hard SVM is as follows:

$$\min_{w,b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{s.t.} \;\; y^{(i)}(\mathbf{w}^\intercal \mathbf{x}^{(i)} + b) \geq 1 \;\; , \forall (x^{(i)}, y^{(i)}) \in \mathcal{D} \tag{1}$$

(a) What is the optimal $\mathbf{w}$ and b? Show all your work and reasoning. (Hint: Draw it out.)

Your answer:

(b) Which of the examples are support vectors?

Your answer:

(c) A standard quadratic program is as follows,

$$\begin{aligned} \underset{\mathbf{z}}{\text{minimize}} \quad & \frac{1}{2}\mathbf{z}^\intercal P \mathbf{z} + \mathbf{q}^\intercal \mathbf{z} \\ \text{subject to} \quad & G\mathbf{z} \leq \mathbf{h} \end{aligned}$$

Rewrite Equation (1) into the above form. ($i.e.$ define $\mathbf{z}, P, \mathbf{q}, G, \mathbf{h}$ using $\mathbf{w}, b$ and values in $\mathcal{D}$). Write the constraints in the **same order** as provided in $\mathcal{D}$ and typeset it using `bmatrix`.

Your answer:

(d) Recall that for a soft-SVM we solve the following optimization problem.

$$\min_{w,b,\xi^{(i)}} \frac{1}{2}\|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^{|D|} \xi^{(i)} \quad \text{s.t.} \;\; y^{(i)}(\mathbf{w}^\intercal \mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)}, \xi^{(i)} \geq 0 \;\; , \forall (x^{(i)}, y^{(i)}) \in \mathcal{D} \tag{2}$$

Describe what happens to the margin when $C = \infty$ and $C = 0$.

Your answer:

2. **[4 points]** Kernels

(a) If $K_1(\mathbf{x}, \mathbf{z})$ and $K_2(\mathbf{x}, \mathbf{z})$ are both valid kernel functions, and α and β are positive, prove that

$$\alpha K_1(\mathbf{x}, \mathbf{z}) + \beta K_2(\mathbf{x}, \mathbf{z})$$

is also a valid kernel function.

Your answer:

(b) Show that $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^\intercal \mathbf{z})^2$ is a valid kernel, for $\mathbf{x}, \mathbf{z} \in \mathbb{R}^2$.
($i.e.$ write out the $\Phi(\cdot)$, such that $K(\mathbf{x}, \mathbf{z}) = \Phi(\mathbf{x})^\intercal \Phi(\mathbf{z})$

Your answer: