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 $y^{(i)} \in \{1, -1\}$

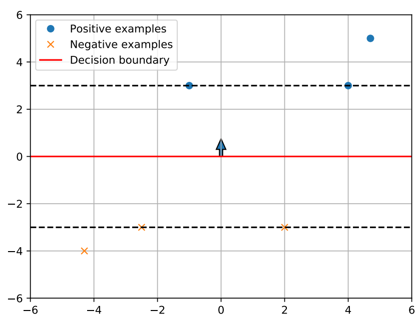
i	$x_1^{(i)}$	$x_2^{(i)}$	$y^{(i)}$
1	-1	3	1
2	-2.5	-3	-1
3	2	-3	-1
4	4.7	5	1
5	4	3	1
6	-4.3	-4	-1

Recall a hard SVM is as follows:

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

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

Solution:



From geometry, it is easy to see that only examples 1,2,3,5 matters. The four points form two parallel lines. To maximize the margin, the center line between the two parallel lines is the decision boundary, therefore $b = 0$. From geometry, the decision boundary has to be a horizontal line, therefore $\mathbf{w} = [0, c]$, for some $c \in \mathbb{R}$. Lastly, we solve for c using the support vectors. For example, using example one,

$$0 \cdot (-1) + 3 \cdot c + 0 = 1$$

, then $c = \frac{1}{3}$.
Finally, $\mathbf{w} = [0, \frac{1}{3}]$, $b = 0$.

(b) Which of the examples are support vectors?

Solution: 1,2,3,5. Easy to see if drawn out.

(c) A standard quadratic program is as follows,

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

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`.

Solution: $\mathbf{z} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ b \end{bmatrix}$ $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $q = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ $G = -\begin{bmatrix} -1 & 3 & 1 \\ 2.5 & 3 & -1 \\ -2 & 3 & -1 \\ 4.7 & 5 & 1 \\ 4 & 3 & 1 \\ 4.3 & 4 & -1 \end{bmatrix}$ $h = -\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

We also accepted general form.

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

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

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

Solution:
When $C \to \infty$ behaves the same as a hard-margin SVM.
When $C \to 0$ margin $\to \infty$, as no penalty on slack variables.

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.

Solution:

K_1 and K_2 are valid kernels, then there exists a corresponding Φ_1, Φ_2, such that $K_1(\mathbf{x}, \mathbf{z}) = \Phi_1(\mathbf{x})^\intercal\Phi_1(\mathbf{z})$ and $K_2(\mathbf{x}, \mathbf{z}) = \Phi_2(\mathbf{x})^\intercal\Phi_2(\mathbf{z})$.
Let $K_3(\mathbf{x}, \mathbf{z}) = \alpha K_1(\mathbf{x}, \mathbf{z}) + \beta K_2(\mathbf{x}, \mathbf{x})$ and $\Phi(x) = \begin{bmatrix} \sqrt{\alpha}\Phi_1(\mathbf{x}) \\ \sqrt{\beta}\Phi_2(\mathbf{x}) \end{bmatrix}$
$K_3(\mathbf{x}, \mathbf{z}) = \Phi(x)^\intercal\Phi(x)$
We also accepted other valid proofs.

(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 $K(\mathbf{x}, \mathbf{z}) = \Phi(\mathbf{x})^\intercal\Phi(\mathbf{z})$)

Solution:

First expand the kernel, $x_1^2 z_1^2 + 2x_1 x_2 z_1 z_2 + x_2^2 z_2^2$

then with some algebra, $\Phi = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1 x_2 \\ x_2^2 \end{bmatrix}$

$K(\mathbf{x}, \mathbf{z}) = \Phi(x)^\intercal \Phi(x)$

We also accepted other valid proofs.