# Homework 4: SVMs (10 points plus 4 bonus points)

1. An SVM is trained using the follow samples:

| sample ID | feature a | feature b | feature c | label |
|---|---|---|---|---|
| 1 | 1.1764 | 4.2409 | 0.9750 | 1 |
| 2 | 1.0400 | 3.8676 | 0.4243 | 1 |
| 3 | 1.0979 | 1.0227 | 0.4484 | 1 |
| 4 | 2.0411 | 4.7610 | 0.6668 | -1 |
| 5 | 2.0144 | 4.1217 | 1.2470 | -1 |
| 6 | 2.1454 | 4.4439 | 0.3974 | -1 |

Suppose (may violate KKT conditions) the $\lambda$'s are sequentially:
$\lambda_1 = 1, \lambda_2 = 0.7383, \lambda_3 = 0, \lambda_4 = 0.0411, \lambda_5 = 1, \lambda_6 = 0.6972,$
what is the corresponding $\mathbf{w}$?
Note that this $\mathbf{w}$ has no bias term.
Be sure to include steps. If you have only the final answer, you won't get any point.

$$\mathbf{w} = \sum_{\mathbf{x}_k \in N_s} \lambda_k y_k \mathbf{x_k}$$

$$= (1)(1)\mathbf{x_1} + (0.7383)(1)\mathbf{x_2} + (0)(1)\mathbf{x_3} + (0.0411)(-1)\mathbf{x_4} + (1)(-1)\mathbf{x_5} + (0.6972)(-1)\mathbf{x_6}$$

$$= [-1.6498, -0.3193, -0.2632]$$

2. Let $w_b$ be $3.3149$. Using the $\mathbf{w}$ obtained above, what is the prediction for a new sample $[1, 1, 0]^T$?

$\mathbf{w}^T\mathbf{x} + w_b = [-1.6498, -0.3193, -0.2632] * [1, 1, 0]^T + 3.3149 = -1.9691 < 0$

```
prediction is: -1
```

3. What are the equations of the two gutters per the $\mathbf{w}$ and $w_b$ obtained above?

$\mathbf{w}^T\mathbf{x} + w_b = d_1$ ==> $\mathbf{w}^T\mathbf{x} + w_b = +1$
$\mathbf{w}^T\mathbf{x} + w_b = -d_2$ ==> $\mathbf{w}^T\mathbf{x} + w_b = -1$

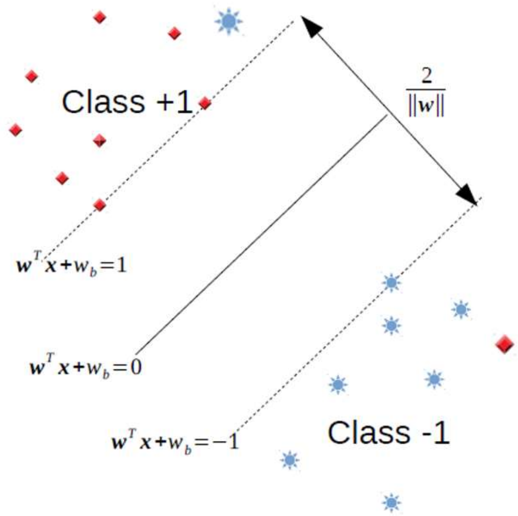
4. With the $\mathbf{w}$ obtained above, and the assumption that $w_b$ is 1, identify samples that fall into the margin and those do not. A sample falls into the margin if it is between the two gutters, i.e.,

$$-1 < \mathbf{w}^T\mathbf{x} + w_b < 1$$

where $\mathbf{x}$ is the (unaugmented) feature vector of the sample.
Show your steps, especially the value of the prediction $\mathbf{w}^T\mathbf{x} + w_b$. If you have only the final answer, you won't get any point.
Please check over all four samples, as the $\lambda$'s above are toy examples and may not satify KKT conditions.

| **ID** | $\mathbf{w}$ | $\mathbf{x}$ | $w_b$ | $\mathbf{w}^T\mathbf{x} + w_b$ | **In Margin?** |
|---|---|---|---|---|---|
| 1 | [-1.6498, -0.3193, -0.2632] | [1.1764, 4.2409, 0.9750] | 1 | -2.5516 | NO |
| 2 | [-1.6498, -0.3193, -0.2632] | [1.0400, 3.8676, 0.4243] | 1 | -2.0625 | NO |
| 3 | [-1.6498, -0.3193, -0.2632] | [1.0979, 1.0227, 0.4484] | 1 | -1.2559 | NO |
| 4 | [-1.6498, -0.3193, -0.2632] | [2.0411, 4.7610, 0.6668] | 1 | -4.0632 | NO |
| 5 | [-1.6498, -0.3193, -0.2632] | [2.0144, 4.1217, 1.2470] | 1 | -3.9678 | NO |
| 6 | [-1.6498, -0.3193, -0.2632] | [2.1454, 4.4439, 0.3974] | 1 | -4.0632 | NO |

5. For an SVM, if a (misclassified) sample $\mathbf{x}_i$ is on the outter side (not the margin side) of the gutter for the opposing class, what conditions below hold? And why? You could use proof-by-contradition to eliminate false choices. (If you do not answer the why part, you get no point.)




In the above case, the misclassified samples are the blue one in the {+1} class and the red one in the {-1} class. In this case:
RED: $\mathbf{w}^T\mathbf{x} + w_b \le -1 | y_i = +1$
BLUE: $\mathbf{w}^T\mathbf{x} + w_b \ge +1 | y_i = -1$

If we multiply the corresponding label $y_i$ to the above cases, in either cases, we get:

$y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \le -1$

In other words, since the sample is misclassified, its label $y_i$ and the prediction $\mathbf{w}^T\mathbf{x}_i + w_b$ will always be oppsited to each other, in our case, always one $\ge$ +1, the other one $\le$ -1. Thus, the multiplication will always be less or equal to -1. So we can conclude the Ture/False as follows (NOTE: I marked TRUE to 4 and 6 because $\le -1$ is within the range of $\le 1$ and $\le 0$):

1. $y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \ge -1$ FALSE.
2. $y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \le -1$ TRUE.
3. $y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \ge 1$ FALSE.
4. $y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \le 1$ TRUE.
5. $y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \ge 0$ FALSE.
6. $y_i(\mathbf{w}^T\mathbf{x}_i + w_b) \le 0$ TRUE.

6. Given a dataset, in cross validation, are the traning sets always the same? What about the test sets?

Not always the same, both traning and test sets are selected dynamically to apply onto the dataset to get a more robust model.