---

(1) (1.0 pt.) In the hard-margin SVM we have the constraints $y_i\mathbf{w}^T\mathbf{x}_i \geq 1$ for all $i = 1, \ldots, m$. For a pair, $(y, \mathbf{x})$, of a new sample example we have $y_i\mathbf{w}^T\mathbf{x}_i < 1$ but $y_i\mathbf{w}^T\mathbf{x}_i > 0$. What does it mean?

1- This example is wrongly classified.

2- This example is classified correctly but it is within the margin. *

3. We can not decide on which class contains this example.

Solution: A point is correctly classified if the signs of the label, y_i, and of the prediction, $\mathbf{w}^T\mathbf{x}_i$, agree, thus $y_i\mathbf{w}^T\mathbf{x}_i > 0$ for both the positive and the negative cases. Here we assume any $\mathbf{x}$ not only the training examples.

(2) (1.0 pt.) In the soft-margin SVM we have the constraints $y_i\mathbf{w}^T\mathbf{x}_i \geq 1 - \xi_i$ for all $i = 1, \ldots, m$. If a pair, $(y, \mathbf{x})$, of a new sample example is classified correctly what could be the value of the corresponding slack variable ξ?

1- $0 \leq \xi < 1$ *

2- $\xi = 0$

3. $\xi < 0$

Solution: An example is correctly classified if $y_i\mathbf{w}^T\mathbf{x}_i > 0$. We have two cases:

- $y_i\mathbf{w}^T\mathbf{x}_i \geq 1$, then $\xi_i = 0$,
- $y_i\mathbf{w}^T\mathbf{x}_i < 1$ and $y_i\mathbf{w}^T\mathbf{x}_i > 0$, then $1 > \xi_i = 1 - y_i\mathbf{w}^T\mathbf{x}_i > 0$,

Finally we have $0 \leq \xi_i < 1$. Be aware, in the second case the strict inequality $\xi_i < 1$ is needed. .

(3) (3.0 pt.) What happens if all labels are positive in an SVM problem? It is called as one-class classification problem. What does the SVM do in this case?

Hint: Think about what the constraints could force in this case.

1- It separates all the examples from the origin. *

2- The result has no any meaning.

3. All slacks will be 0.

Solution: Even if all examples are positive the SVM can process them. It can start from a feasible $\mathbf{w} \neq 0$ then for any $\mathbf{x}_i$ of the training there is a $\xi_i \geq 0$ such that $\mathbf{w}^T\mathbf{x}_i \geq 1 - \xi_i$ where the y_i is dropped since it is equal to 1. Then the SVM finds the largest optimal margin at the smallest loss and provides the optimal $\mathbf{w}$ and ξ_is. In the optimum the majority of the examples have to be above the hyperplane $\mathbf{w}^T\mathbf{x}_i = 1$ otherwise the solution is not optimal. Now if the origin $\mathbf{x}_0 = \mathbf{0}$ is substituted into the constraints we have $\mathbf{w}^T\mathbf{x}_0 = 0$, thus it is on the hyperplane defined $\mathbf{w}^T\mathbf{x}_0 = 0$, and below the hyperplane $\mathbf{w}^T\mathbf{x}_i = 1$. The distance between the two hyperplanes is exactly the margin which is maximized by the SVM, therefore those positive examples are separated from the origin when the margin is maximized.

(4) (5.0 pt.)

Let 6 points be given in the plane, $X = \{(0,1), (2,1), (2,-1), (0,-1), (-2,-1), (-2,1)\}$, and the corresponding labels in the same order $y = \{1, 1, 1, -1, -1, -1\}$. Compute $\mathbf{w}$ by the algorithm given by Slide 20 of Lecture 7 by taking the examples in the given fix order instead of randomly drawing them. Let $\lambda = 1$, and the learning speed $\eta = 0.1$. Which of these coordinates can give the solution for $\mathbf{w}$? Round the numbers up to 2 decimals, and take the closest one.

1- (0.54, -0.07)

2. (0.66, 0.12) *

3- (-0.66, -0.12)

Solution: In Python the code it looks like this:

```
## ####################################################
import numpy as np

## Load the data
X=np.array([(0,1),(2,1),(2,-1),(0,-1),(-2,-1),(-2,1)])
y=np.array([1,1,1,-1,-1,-1])

m,n=X.shape

## initialize w, eta, lambda
w=np.zeros(n)
eta=0.1
xlambda=1.0
for i in range(m):              ## process all sample examples
  if y[i]*np.dot(w,X[i])<1:    ## check the conditions
    J=-y[i]*X[i]+xlambda*w     ## compute gradient
  else:
    J=xlambda*w                ## compute gradient
  w=w-eta*J                    ## update w

  print(w)

print(w)
## ####################################################
```