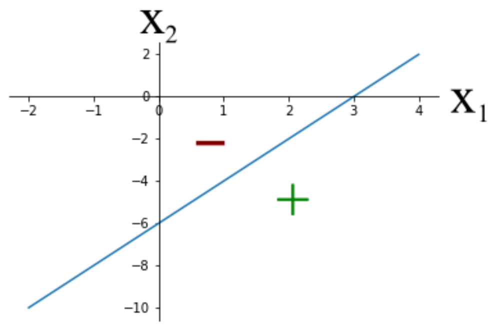
DSE 220: Machine learning

Worksheet 9 — Solutions

1. The decision boundary plot should look something like the plot below.



2. (a) **Definitely true.** If the data set were not linearly separable, Perceptron would never converge.
   (b) **Definitely true.** Since the data is linearly separable, Perceptron is guaranteed to converge, no matter what the ordering of the points might be.
   (c) **Possibly false.** Different orderings of the data can produce different numbers of updates before convergence. We saw examples of this in class.
   (d) **Possibly false.** There could be several updates on any given data point, and thus k is not necessarily upper-bounded by n.

3. Each time the Perceptron algorithm performs an update a point with label y, it updates its offset b as $b = b + y$. Thus if we start with $b = 0$ and perform p updates on points with $y = -1$ and q updates on points with $y = +1$, then the final value of b is $b = q - p$.

4. *Perceptron project.*
   (a) The classification code can be written as follows.

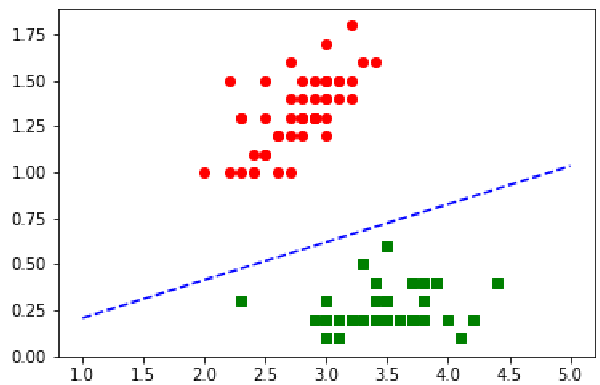
```
def classify(w, b, x):
    return np.sign(np.dot(w,x) + b)
```

The perceptron algorithm can be written as follows.

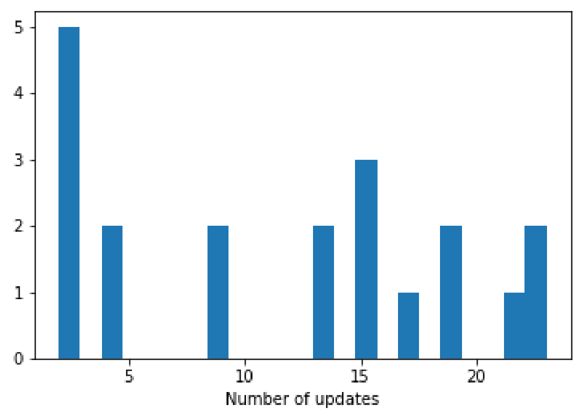
```
def perceptron(data, labels):
    n = len(labels)
    inds = np.random.permutation(n)
    data = data[inds,:]
    labels = labels[inds]
    n_correct = 0
    w = np.zeros(np.shape(data)[1])
```

```
    b = 0
    while(n_correct < n):
        n_correct = 0
        for i in range(n):
            if (classify(w, b, data[i,:]) == labels[i]):
                n_correct += 1
            else:
                w = w + labels[i]*data[i,:]
                b = b + labels[i]
    return(w,b)
```

(c) The perceptron boundary should look something like the following plot.



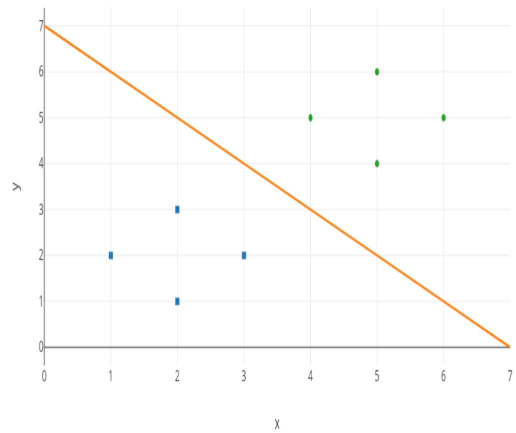
(d) The histogram should look something like the following.



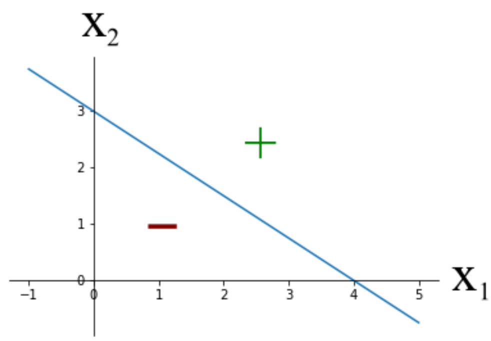
5. (a) Decision boundary is shown in figure.

(b) The margin is $\sqrt{2}$.

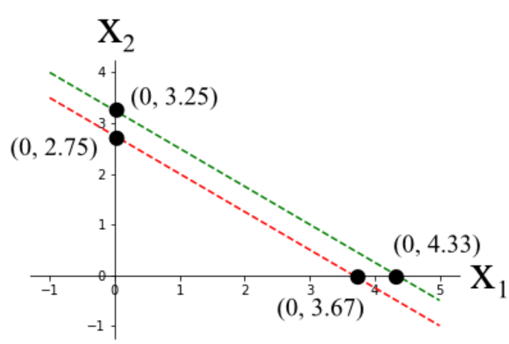
(c) w lies in direction of $\begin{pmatrix}1\\1\end{pmatrix}$ and has length $1/\sqrt{2}$ (since the margin is $\sqrt{2}$; therefore, $w = \begin{pmatrix}1/2\\1/2\end{pmatrix}$. We know that point $x_0 = (4,3)$ lies on decision boundary; solving $w \cdot x_0 + b = 0$ yields $b = -7/2$.



6. (a) The decision boundary plot should look something like the plot below.



(b) The left- and right-hand boundary plot should look something the plot below.



(c) The margin of this classifier is
$$\gamma = \frac{1}{\|w\|} = \frac{1}{\sqrt{3^2+4^2}} = \frac{1}{5}.$$

(d) The point $x = (2, 2)$ satisfies
$$w \cdot x + b = 6 + 8 - 12 = 2 > 0$$

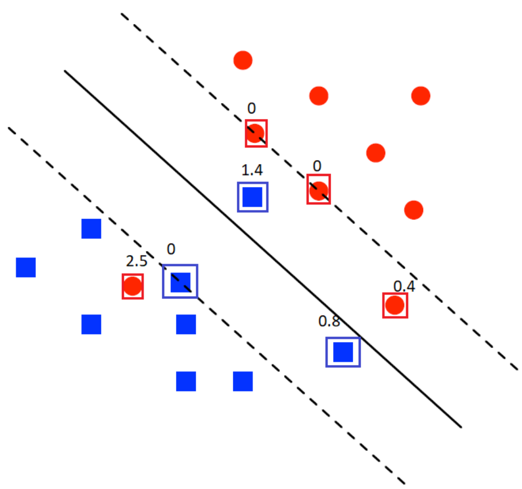
Thus this point would be classified as +1.

(e) The support vectors are the points x such that $x \cdot w + b = \pm 1$. We are told that there are two distinct support vectors $x^{(1)}, x^{(2)} \in \mathbb{R}^2$ and they both satisfy $x_1^{(1)} = 1 = x_1^{(2)}$. Then it must be the case that $x^{(1)} \cdot w + b = +1$ and $x^{(2)} \cdot w + b = +1$. Solving for $x_2^{(1)}$ gives us
$$1 = x^{(1)} \cdot w + b = 3 + 4x_2^{(1)} - 12 = 4x_2^{(1)} - 9.$$

Thus $x^{(1)} = (1, 2.5)$. Solving for $x_2^{(2)}$ gives us
$$-1 = x^{(2)} \cdot w + b = 3 + 4x_2^{(2)} - 12 = 4x_2^{(2)} - 9.$$

Thus $x^{(2)} = (1, 2)$.

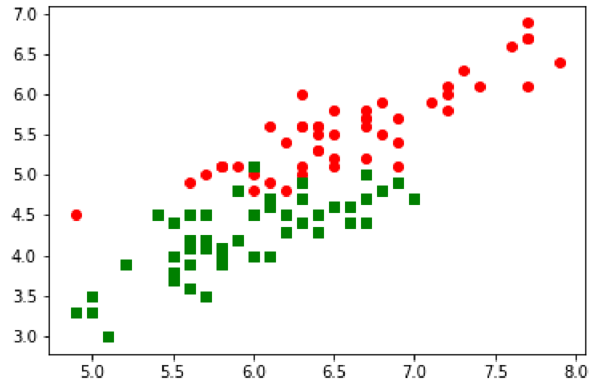
7. (a) Support vectors and their respective slack variables are marked in figure.



(b) The margin decreases if the factor C is increased.

8. *Support vector machine.*

(a) The data is not linearly separable. We can see this by inspecting the scatter plot.



(b) The table you produce should look something like the following.

C value	1.5	3.0	4.5	6.0	7.5	9.0	10.5	12.0	13.5	15.0
Training error	0.07	0.05	0.06	0.05	0.05	0.05	0.05	0.04	0.05	0.07
# of support vectors	27	22	21	19	19	19	18	17	16	16

(c) The boundary plot should look something like the following.