CMSE 820 Homework 8

Wei-Chien Liao (liaowei2@msu.edu)

1. Let w^* be the unit vector in the definition of linear separability, then

$$y_j\left((w^*)^T x^{(j)}\right) > \gamma > 0. \tag{1}$$

Compute the projection $w^{(k)}$ on w^*. Assume at this stage (x^i, y_i) is misclassified. Then

$$\begin{aligned}
&w^{(k)} = w^{(k-1)} + \eta(y_i - \text{sign}((w^{(k-1)})^T x^{(i)}))x^{(i)} \\
\Rightarrow &(w^{(k)})^T w^* = (w^{(k-1)})^T w^* + \eta(y_i - \text{sign}((w^{(k-1)})^T x^{(i)}))((x^{(i)})^T w^*)
\end{aligned}$$

Note that since $y_i((w^{(k-1)})^T x^{(i)}) < 0$, $-\text{sign}((w^{(k-1)})^T x^{(i)}) = y_i$. Using this, we have

$$\begin{aligned}
(w^{(k)})^T w^* &= (w^{(k-1)})^T w^* + 2\eta y_i((x^{(i)})^T w^*) \\
&> (w^{(k-1)})^T w^* + 2\eta\gamma \quad \text{(by (1))}
\end{aligned}$$

This implies

$$\begin{aligned}
&\sum_{n=1}^{k}(w^{(n)})^T w^* > \sum_{n=1}^{k}((w^{(n-1)})^T w^* + 2\eta\gamma) \\
\Rightarrow &(w^{(k)})^T w^* > (\underbrace{w^{(0)}}_{=0})^T w^* + 2k\eta\gamma = 2k\eta\gamma \\
\Rightarrow &||w^{(k)}|| = ||w^{(k)}||||w^*|| \geq |(w^{(k)})^T w^*| = (w^{(k)})^T w^* > 2k\eta\gamma \\
\Rightarrow &||w^{(k)}||^2 \geq 4k^2\eta^2\gamma^2
\end{aligned} \tag{2}$$

On the other hand,

$$
\begin{aligned}
||w^{(k)}||^2 &= ||w^{(k-1)} + \eta(y_i - \text{sign}((w^{(k-1)})^T x^{(i)}))x^{(i)}||^2 \\
&= ||w^{(k-1)} + 2\eta y_i x^{(i)}||^2 \\
&= \left(w^{(k-1)} + 2\eta y_i x^{(i)}, w^{(k-1)} + 2\eta y_i x^{(i)}\right) \\
&= ||w^{(k-1)}||^2 + ||2\eta y_i x^{(i)}||^2 + 4\eta y_i \left(w^{(k-1)}, x^{(i)}\right) \\
&= ||w^{(k-1)}||^2 + 4\eta^2 ||x^{(i)}||^2 + 4\eta \underbrace{y_i \left((w^{(k-1)})^T x^{(i)}\right)}_{<0} \\
&\leq ||w^{(k-1)}||^2 + 4\eta^2 R^2 \quad (\because ||x^{(i)}|| \leq R)
\end{aligned}
$$

This implies

$$
\begin{aligned}
&\sum_{n=1}^{k} ||w^{(n)}||^2 \leq \sum_{n=1}^{k} \left(||w^{(n-1)}||^2 + 4\eta^2 R^2\right) \\
\Rightarrow &||w^{(k)}||^2 \leq ||\underbrace{w^{(0)}}_{=0}||^2 + 4k\eta^2 R^2 \\
\Rightarrow &||w^{(k)}||^2 \leq 4k\eta^2 R^2 \qquad (3)
\end{aligned}
$$

Combining (2) and (3) gives,

$$
\begin{aligned}
&4k^2\eta^2\gamma^2 \leq ||w^{(k)}||^2 \leq 4k\eta^2 R^2 \\
\Rightarrow &4k^2\eta^2\gamma^2 \leq 4k\eta^2 R^2 \\
\Rightarrow &k \leq \frac{R^2}{\gamma^2}.
\end{aligned}
$$

□

2. (a) Let

$$
u^{(1)} = \begin{bmatrix} u_1^{(1)} \\ \vdots \\ u_n^{(1)} \end{bmatrix}, u^{(2)} = \begin{bmatrix} u_1^{(2)} \\ \vdots \\ u_n^{(2)} \end{bmatrix} \in \mathbb{R}^n_{\geq 0}
$$

and

$$
v^{(1)} = \begin{bmatrix} v_1^{(1)} \\ \vdots \\ v_r^{(1)} \end{bmatrix}, u^{(2)} = \begin{bmatrix} v_1^{(2)} \\ \vdots \\ v_r^{(2)} \end{bmatrix} \in \mathbb{R}^r.
$$

Let $(u^{(3)}, v^{(3)}) = t(u^{(1)}, v^{(1)}) + (1-t)(u^{(2)}, v^{(2)})$ for some $t \in (0,1)$. Define

$$
L(x_0, (u, v)) := L(x_0, u, v).
$$

Then

$$
\begin{aligned}
&L(x_0, (u^{(3)}, v^{(3)}))\\
=&L(x_0, t(u^{(1)}, v^{(1)}) + (1-t)(u^{(2)}, v^{(2)}))\\
=&L(x_0, (tu^{(1)} + (1-t)u^{(2)}, tv^{(1)} + (1-t)v^{(2)}))\\
=&L(x_0, tu^{(1)} + (1-t)u^{(2)}, tv^{(1)} + (1-t)v^{(2)})\\
=&f(x_0) + \sum_{i=1}^{n} \left[tu_i^{(1)} + (1-t)u_i^{(2)}\right] h_i(x_0) + \sum_{j=1}^{r} \left[tv_j^{(1)} + (1-t)v_j^{(2)}\right] g_j(x_0)\\
=&t\left[f(x_0) + \sum_{i=1}^{n} u_i^{(1)} h_i(x_0) + \sum_{j=1}^{r} v_j^{(1)} g_j(x_0)\right]\\
&+ (1-t)\left[f(x_0) + \sum_{i=1}^{n} u_i^{(2)} h_i(x_0) + \sum_{j=1}^{r} v_j^{(2)} g_j(x_0)\right]\\
=&tL(x_0, u^{(1)}, v^{(1)}) + (1-t)L(x_0, u^{(2)}, v^{(2)})\\
=&tL(x_0, (u^{(1)}, v^{(1)})) + (1-t)L(x_0, (u^{(2)}, v^{(2)}))
\end{aligned}
$$

Therefore, $L(x_0, u, v)$, as a function of (u, v), is a convex function. $\square$

(b) $x_0 \in C \Rightarrow h_i(x_0) \leq 0,\ i = 1, \cdots, n$ and $g_j(x_0) = 0,\ j = 1, \cdots, r$. Therefore,

$$
\begin{aligned}
L(x_0, u, v) &= f(x_0) + \sum_{i=1}^{n} u_i h_i(x_0) + \sum_{j=1}^{r} v_j g_j(x_0)\\
&= f(x_0) + \sum_{i=1}^{n} u_i h_i(x_0).
\end{aligned}
$$

Note that $u_i(x_0) \geq 0$ and $h_i(x_0) \leq 0$ implies $u_i h_i(x_0) \leq 0$ for $i = 1, \cdots, n$, so

$$\sum_{i=1}^{n} u_i h_i(x_0) \leq 0.$$

Then,

$$L(x_0, u, v) = f(x_0) + \sum_{i=1}^{n} u_i h_i(x_0) \leq f(x_0).$$

Therefore,

$$\begin{aligned}
\max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} L(x_0,u,v) &= \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r}\left[f(x_0)+\sum_{i=1}^n u_i h_i(x_0)\right] \\
&= f(x_0) + \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} \sum_{i=1}^n u_i h_i(x_0) \\
&= f(x_0).
\end{aligned}$$

□

(c) $x_0 \notin C \Rightarrow h_i(x_0) > 0,\ i = 1,\cdots,n$ or $g_j(x_0) \neq 0,\ j = 1,\cdots,r$.
Let

$$u_i^{(\ell)} = \begin{cases} i\cdot 2^\ell, & \text{if } h_i(x_0) > 0 \\ 0, & \text{if } h_i(x_0) \leq 0 \end{cases} \qquad \text{for } i = 1,\cdots,n,$$

and

$$v_j^{(\ell)} = \operatorname{sign}(g_j(x_0))\cdot j\cdot 2^\ell \qquad \text{for } j = 1,\cdots,r.$$

Then $\{u^{(\ell)}\} \subset \mathbb{R}^n_{\geq 0}$ and $\{v^{(\ell)}\} \subset \mathbb{R}^r$. Now

$$\begin{aligned}
L(x_0,u^{(\ell)},v^{(\ell)}) &= f(x_0) + \sum_{i=1}^n u_i^{(\ell)} h_i(x_0) + \sum_{j=1}^r \operatorname{sign}(g_j(x_0))\cdot j\cdot 2^\ell g_j(x_0) \\
&= f(x_0) + \sum_{i=1}^n u_i^{(\ell)} h_i(x_0) + \sum_{j=1}^r |g_j(x_0)|\cdot j\cdot 2^\ell
\end{aligned}$$

From the choice of $u^{(\ell)}$, we have $u_i^{(\ell)} h_i(x_0) \geq 0$ for $i = 1,\cdots,n$. This implies,

$$\begin{aligned}
L(x_0,u^{(\ell)},v^{(\ell)}) &= f(x_0) + \sum_{i=1}^n u_i^{(\ell)} h_i(x_0) + \sum_{j=1}^r |g_j(x_0)|\cdot j\cdot 2^\ell \\
&\geq f(x_0) + \sum_{j=1}^r |g_j(x_0)|\cdot j\cdot 2^\ell \to \infty \text{ as } \ell\to\infty
\end{aligned}$$

Therefore, $L(x_0,u^{(\ell)},v^{(\ell)}) \to \infty$ as $\ell \to \infty$.

3. (a)

$$\begin{aligned}
\max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} \min_{x\in\mathbb{R}^p} L(x,u,v) &= \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r}\left[\min_{x\in\mathbb{R}^p} L(x,u,v)\right] \\
&= \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} \ell(u,v) \\
&= \ell^*
\end{aligned}$$

□

(b) Note that

$$-\ell(u,v) = -\min_{x\in\mathbb{R}^p} L(x,u,v) = \max_{x\in\mathbb{R}^p} -L(x,u,v).$$

Let $(u^{(1)},v^{(1)}),(u^{(2)},v^{(2)}) \in \mathbb{R}^n_{\geq 0} \times \mathbb{R}^r$.
Let $(u^{(3)},v^{(3)}) = t(u^{(1)},v^{(1)}) + (1-t)(u^{(2)},v^{(2)})$ for some $t \in (0,1)$.
Then

$$\begin{aligned}
&-\ell(u^{(3)},v^{(3)})\\
=&\max_{x\in\mathbb{R}^p} -L(x,u^{(3)},v^{(3)})\\
=&\max_{x\in\mathbb{R}^p} -\left[f(x) + \sum_{i=1}^n u_i^{(3)} h_i(x) + \sum_{j=1}^r v_j^{(3)} g_j(x)\right]\\
=&\max_{x\in\mathbb{R}^p} \left\{-t\left[f(x) + \sum_{i=1}^n u_i^{(1)} h_i(x) + \sum_{j=1}^r v_j^{(1)} g_j(x)\right]\right.\\
&\left.-(1-t)\left[f(x) + \sum_{i=1}^n u_i^{(2)} h_i(x) + \sum_{j=1}^r v_j^{(2)} g_j(x)\right]\right\}\\
=&\max_{x\in\mathbb{R}^p} -tL(x,u^{(1)},v^{(1)}) - (1-t)L(x,u^{(2)},v^{(2)})\\
\leq&\max_{x\in\mathbb{R}^p} -tL(x,u^{(1)},v^{(1)}) + \max_{x\in\mathbb{R}^p} -(1-t)L(x,u^{(2)},v^{(2)})\\
=&t\max_{x\in\mathbb{R}^p} -L(x,u^{(1)},v^{(1)}) + (1-t)\max_{x\in\mathbb{R}^p} -L(x,u^{(2)},v^{(2)})\\
=&t\left(-\ell(u^{(1)},v^{(1)})\right) + (1-t)\left(-\ell(u^{(2)},v^{(2)})\right)
\end{aligned}$$

This implies $-\ell(u,v)$ is convex, so $\ell(u,v)$ is concave. □

(c) By definition,

$$\begin{aligned}
f^* &= \min_{x\in\mathbb{R}^p} \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} L(x,u,v),\\
\ell^* &= \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} \ell(u,v).
\end{aligned}$$

Therefore,

$$f^* = \min_{x\in\mathbb{R}^p} \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} L(x,u,v) \geq \min_{x\in\mathbb{R}^p} L(x,u,v) = \ell(u,v).$$

Taking maximum on both sides of the inequality gives

$$f^* = \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} f^* \geq \max_{u\in\mathbb{R}^n_{\geq 0}, v\in\mathbb{R}^r} \ell(u,v) = \ell^*.$$

□

(d) Let

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}, v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Then the primal Lagrangian for the given model is

$$L(w, v) = f(w) + \sum_{i=1}^{n} v_i h_i(w) = ||w||^2 + \sum_{i=1}^{n} v_i \left[1 - y_i(w^T x^{(i)})\right].$$

Now derive the dual problem:

$$\begin{aligned}
\ell(v) &= \min_{w \in \mathbb{R}^p} L(w, v) \\
&= \min_{w \in \mathbb{R}^p} w^T w + \sum_{i=1}^{n} v_i \left[1 - y_i(w^T x^{(i)})\right] \\
&= \min_{w \in \mathbb{R}^p} \sum_{j=1}^{p} w_j^2 + \sum_{i=1}^{n} v_i \left[1 - y_i(\sum_{j=1}^{p} w_j x_j^{(i)})\right]
\end{aligned}$$

Computing the partial derivatives of L corresponding to w_k and setting it to zero gives

$$\begin{aligned}
\frac{\partial L}{\partial w_k} &= 2w_k + \sum_{i=1}^{n} \left(-v_i y_i x_k^{(i)}\right) = 0 \\
\Rightarrow w_k &= \frac{1}{2} \sum_{i=1}^{n} v_i y_i x_k^{(i)}, \quad k = 1, \cdots, p
\end{aligned}$$

Define

$$X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(n)} \\ | & | & & | \end{bmatrix}_{p \times n}, \quad Y = \begin{bmatrix} y_1 & & & \\ & y_2 & & \\ & & \ddots & \\ & & & y_n \end{bmatrix}_{n \times n}$$

Then the partial derivatives of L corresponding to w_k can be written as

$$w_k = \frac{1}{2} (XYv)_k, \quad k = 1, \cdots, p,$$

or,

$$w = \frac{1}{2} XYv.$$

Expanding the second term of $L(w, v)$ gives

$$\begin{aligned}\sum_{i=1}^{n} v_i \left[1 - y_i(w^T x^{(i)})\right] &= \sum_{i=1}^{n} v_i \left[1 - y_i(\sum_{j=1}^{p} w_j x_j^{(i)})\right] \\ &= \sum_{i=1}^{n} v_i - \sum_{i=1}^{n}\sum_{j=1}^{p} v_i y_i w_j x_j^{(i)}.\end{aligned}$$

The second term in the above equation can be written as the following matrix-vector multiplications

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{p} v_i y_i w_j x_j^{(i)} \\ =&\underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{1\times n} \underbrace{\begin{bmatrix} y_1 & & \\ & \ddots & \\ & & y_n \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} — & (x^{(1)})^T & — \\ & \vdots & \\ — & (x^{(n)})^T & — \end{bmatrix}}_{n\times p} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_p \end{bmatrix}}_{p\times 1} \\ =&v^T Y X^T w \\ =&v^T Y X^T \left(\frac{1}{2} X Y v\right) \\ =&\frac{1}{2} v^T Y X^T X Y v\end{aligned}$$

Combining all results together, we obtain

$$\begin{aligned}\ell(v) &= \min_{w\in\mathbb{R}^p} L(w, v) \\ &= \left(\frac{1}{2} X Y v\right)^T \left(\frac{1}{2} X Y v\right) + \sum_{i=1}^{n} v_i - \frac{1}{2} v^T Y X^T X Y v \\ &= \frac{1}{4} v^T Y X^T X Y v - \frac{1}{2} v^T Y X^T X Y v + \sum_{i=1}^{n} v_i \\ &= -\frac{1}{4} v^T Y X^T X Y v + v^T \mathbf{1},\end{aligned}$$

where $\mathbf{1}$ is a n-by-1 vector with all entries equal to 1. $\square$

HW8-P4-Perceptron

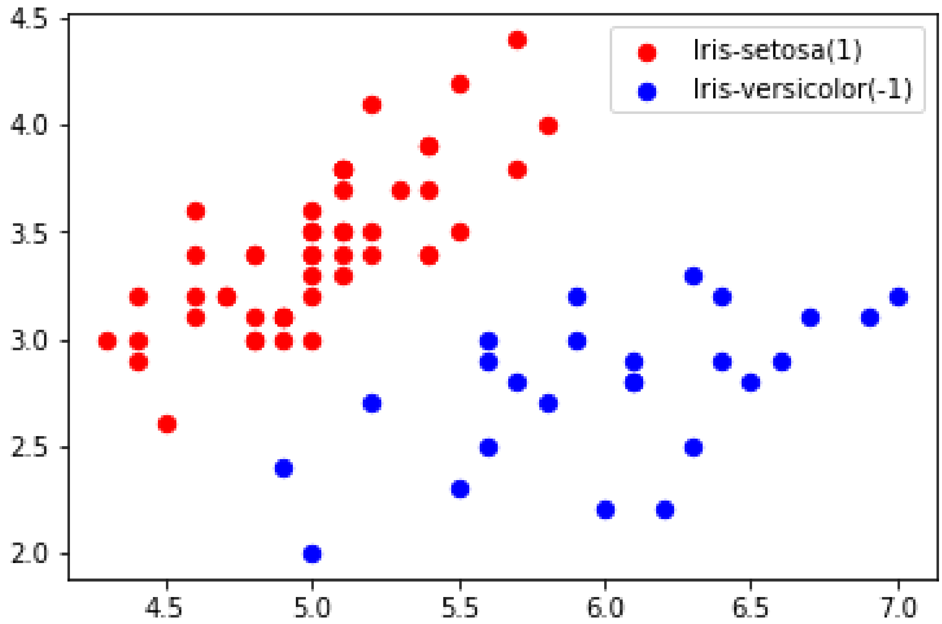
[1]:
```python
# setup libraries
import numpy as np
import matplotlib.pyplot as plt
import sklearn as skl
import matplotlib.pyplot as plt
from matplotlib import cm
import pandas as pd
from sklearn.decomposition import KernelPCA
from sklearn.datasets import make_moons
from scipy.spatial.distance import pdist, squareform
from scipy import exp
import csv
```

[155]:
```python
# Perceptron function
def perceptron(X,y):

    k = 0
    w = np.zeros((X.shape[1],1))
    w = w + np.reshape(y[0]*X[0,:],(X.shape[1],1))
    Xw = np.reshape(X.dot(w),(X.shape[0],1))
    cond = any(Xw[:,0]*y<0)
    print(np.sum(Xw[:,0]*y<0))
    print(cond)
    while cond:
        ind = np.argmax(Xw[:,0]*y<0)
        w = w + np.reshape(y[ind]*X[ind,:],(X.shape[1],1))
        Xw = np.reshape(X.dot(w),(X.shape[0],1))
        cond = any(Xw[:,0]*y<0)
        print(f'In step {k+1}, number of errors = {np.sum(Xw[:,0]*y<0)}')
        k = k+1

    return w,k
```

[193]:
```python
# open and read csv file
sample_data = []
with open('iris.csv') as csvFile:
    datareader = csv.reader(csvFile)
    sample_data_header = next(datareader) # read the titles in first row
```

```
    for row in datareader: # save data to the list object sample_data
        sample_data.append(row)

# Initialize sample matrix X and label vector y
# print(len(sample_data))
X = np.zeros((len(sample_data),2))
y = np.zeros((len(sample_data),1))

# data preprocessing
for k in range(len(sample_data)):
    for n in range(3):
        if n <= 1:
            X[k,n] = sample_data[k][n]
        elif n == 2:
            if sample_data[k][n] == 'Iris-setosa':
                y[k,0] = 1
            elif sample_data[k][n] == 'Iris-versicolor':
                y[k,0] = -1

# setup training and testing data
X_train= X[0:75,:]
y_train= y[0:75,0]
X_test = X[75:,:]

# plot training data
c0s = plt.scatter(X_train[0:50,0],X_train[0:50,1],s=40.
 ↪0,c='r',label='Iris-setosa(1)')
c1s = plt.scatter(X_train[50:75,0],X_train[50:75,1],s=40.
 ↪0,c='b',label='Iris-versicolor(-1)')
plt.legend(fontsize=10,loc=1)
plt.show()


# run perceptron with X_train and y_train
w,iternum = perceptron(X_train,y_train)
print(f'number of iterations: {iternum}')
print(f'weight: {w}')
print(f'weight length: {np.linalg.norm(w)}')

# compute projections to the hyperplane
dotprods = X_train.dot(w)
projs = dotprods/np.linalg.norm(w)
projs_pts = np.mat(np.zeros(X_train.shape))
projs_pts[:,0] = X_train[:,0].reshape((X_train.shape[0],1))-projs*(-w[1,0])
projs_pts[:,1] = X_train[:,1].reshape((X_train.shape[0],1))-projs*w[0,0]
pts = projs_pts[np.array([0,-1]),:]
```

```
# plot results using training data
plt.scatter(X_train[0:50,0],X_train[0:50,1],s=40.0,c='r',label='Iris-setosa(1)')
plt.scatter(X_train[50:75,0],X_train[50:75,1],s=40.
↪0,c='b',label='Iris-versicolor(-1)')
plt.plot(pts[:,0],pts[:,1])
plt.legend(fontsize=10,loc=1)
plt.show()

# plot results using test data
plt.scatter(X_test[0:25,0],X_test[0:25,1],s=40.0,c='b',label='Iris-setosa(-1)')
plt.plot(pts[:,0],pts[:,1])
plt.legend(fontsize=10,loc=1)
plt.show()
```




```
(75, 1)
(75,)
25
True
In step 1, number of errors = 50
In step 2, number of errors = 25
In step 3, number of errors = 50
In step 4, number of errors = 25
In step 5, number of errors = 50
In step 6, number of errors = 25
In step 7, number of errors = 50
In step 8, number of errors = 2
```

```
In step 9, number of errors = 50
In step 10, number of errors = 0
number of iterations: 10
weight: [[-3. ]
 [ 5.2]]
weight length: 6.003332407921454
```