Solution to Homework 05

Name: Chen Shen

NetID: cs5236

1.

(a) First, let

$$\mathbf{A} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \vdots & \cdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} \end{bmatrix}.$$

Then we have $\mathbf{z} = \mathbf{A}\mathbf{w}$, where

$$z_i = w_0 + \sum_{j=1}^{d} w_j x_{ij}.$$

So if we let

$$g(\mathbf{z}) = \sum_{i=1}^{n} g_i(z_i),$$

where

$$g_i(z_i) = \left[y_i - \frac{1}{z_i}\right]^2$$

Then we will have $J(\mathbf{w}) = g(\mathbf{A}\mathbf{w})$.

(b) The gradient of $g(\mathbf{z})$ is

$$\nabla_{\mathbf{z}} g(\mathbf{z}) = \left[g_1'(z_1), \quad \cdots, \quad g_n'(z_n)\right]^\top,$$

where

$$g_i'(z_i) = -\frac{1}{z_i^2}\left[y_i - (-\frac{1}{z_i})\right]$$

Based on the forward-backward rule,

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \mathbf{A}^\top \nabla_{\mathbf{z}} g(\mathbf{z}),$$

where

$$\mathbf{z} = \mathbf{A}\mathbf{w}$$

(c)

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \alpha \nabla_{\mathbf{w}} f(\mathbf{w}^k)$$

(d)

```
n = X.shape[0]
A = np.column.stack((np.ones(n), X))
z = A.dot(w)
yerr = y - 1 / z
J = np.sum(yerr ** 2)

ggrad = -yeer / (z ** 2)
Jgrad = A.T.dot(ggrad)
```

2.

(a)

$$\nabla J(\mathbf{w}) = \begin{bmatrix} \frac{\partial J}{\partial w_1} & \frac{\partial J}{\partial w_2} \end{bmatrix}^\top = \begin{bmatrix} b_1 w_1 & b_2 w_2 \end{bmatrix}^\top$$

(b)

$$\mathbf{w}^* = 0$$

(c)

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \alpha \nabla J(\mathbf{w}^k)$$
$$\Rightarrow wi^{k+1} = wi^k \alpha b_i w_i k = \rho_i w_i^k$$

(d) In order to obtain $\mathbf{w}^k \to \mathbf{w}^*$, we should have

$$|1 - b_i \alpha| < 1$$
$$\Rightarrow \alpha < \frac{2}{b_i}$$

where $i = 1, 2$.

(e) For $\alpha = 2/(b_1 + b_2)$, we have

$$\rho_1 = 1 - b_1 \alpha = \frac{b_2 - b_1}{b_2 + b_1}, \rho_2 = 1 - b_2 \alpha = \frac{b_1 - b_2}{b_1 + b_2}$$

Let $C = \frac{b_2 - b_1}{b_2 + b_1}$, then we have $|\rho_i| = C$ $(i = 1, 2)$.
Since $K = \frac{b_2}{b_1}$, there is $C = \frac{K-1}{K+1}$.
Based on the previous problems, $w_i^k = \rho_i^k w_i^0$, i.e. $|w_i^k| = C|w_i^0|$.
Thus,

$$||w^k||^2 = |w_1^k|^2 + |w_2^k|^2 = C^{2k} \left[|w_1^0|^2 + |w_2^0|^2 \right] = C^{2k} ||\mathbf{w}^0||^2$$

3.

(a)

$$z_i = \mathbf{x}_i^\top \mathbf{P} \mathbf{x}_i = \sum_{j,k} x_{ij} x_{ik} P_{jk}$$

So

$$\frac{\partial z_i}{\partial P_{jk}} = x_{ij} x_{ik}$$

Thus,

$$\nabla_{\mathbf{P}} z_i = [x_{ij} x_{ik}] = \mathbf{x}_i \mathbf{x}_i^\top$$

(b)

$$\nabla_{\mathbf{P}} J(\mathbf{P}) = \nabla_{z_i} J \nabla_{\mathbf{P}} z_i = \sum_{i=1}^n \left[\frac{1}{y_i} - \frac{1}{z_i} \right] \mathbf{x}_i \mathbf{x}_i^\top$$

(c)

```python
n = X.shape[0]
z = np.zeros(n)
for i in range(n):
    z[i] = np.dot(X[i,:], np.dot(P, X[i,:]))

J = np.sum(z/y - np.log(z))
g = 1/y - 1/z

Jgrad = np.zeros((n,n))
for i in range(n):
    xi = X[i,:]
    Jgrad += g[i] * xi[:,None] * xi[None,:]
```

(d)

```python
n = X.shape[0]
z = np.sum(XP*X, axis=1)

J = np.sum(z/y - np.log(z))
g = 1/y - 1/z

GX = g[:,None] * X
Jgrad = np.dot(X.T, GX)
```

4.

(a) First, we have

$$J_1(\mathbf{w}_1) = J(\mathbf{w}_1, \hat{\mathbf{w}}_2)$$

$$\begin{aligned}\frac{\partial J_1}{\partial w_{1j}} &= \frac{\partial J(\mathbf{w}_1, \hat{\mathbf{w}}_2)}{\partial w_{1j}} \\ &= \left.\frac{\partial J(\mathbf{w}_1, \mathbf{w}_2)}{\partial w_{1j}}\right|_{\mathbf{w}_2=\hat{\mathbf{w}}_2} + \sum_k \left.\frac{\partial J(\mathbf{w}_1, \mathbf{w}_2)}{\partial w_{2k}}\right|_{\mathbf{w}_2=\hat{\mathbf{w}}_2} \frac{\partial w_{2k}}{\partial w_{1j}}\end{aligned}$$

Also we have

$$\nabla_{\mathbf{w}_2} J(\mathbf{w}_1, \mathbf{w}_2)|_{\mathbf{w}_2=\hat{\mathbf{w}}_2} = 0$$

Thus,

$$\left.\frac{\partial J(\mathbf{w}_1, \mathbf{w}_2)}{\partial w_{2k}}\right|_{\mathbf{w}_2=\hat{\mathbf{w}}_2} = 0$$

So

$$\frac{\partial J_1}{\partial w_{1j}} = \left.\frac{\partial J(\mathbf{w}_1, \mathbf{w}_2)}{\partial w_{1j}}\right|_{\mathbf{w}_2=\hat{\mathbf{w}}_2}$$

In conclusion,

$$\nabla_{\mathbf{w}_1} J_1(\mathbf{w}_1) = \nabla_{\mathbf{w}_1} J(\mathbf{w}_1, \mathbf{w}_2)|_{\mathbf{w}_2=\hat{\mathbf{w}}_2}$$

(b) In this problem, we can treat $\mathbf{a}$ as a constant.
Let

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 & \cdots & \hat{y}_n \end{bmatrix}^\top,$$

where

$$\hat{y}_i = \sum_{j=1}^{d} b_j e^{-a_j x_i},$$

and

$$\mathbf{A} = \begin{bmatrix} e^{-a_1 x_1} & \cdots & e^{-a_d x_1} \\ \vdots & \cdots & \vdots \\ e^{-a_1 x_n} & \cdots & e^{-a_d x_n} \end{bmatrix}.$$

Also we have $\hat{\mathbf{y}} = \mathbf{A}\mathbf{b}$.
Thus,

$$\hat{\mathbf{b}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}.$$

(c)

$$\nabla_{\mathbf{a}} J(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} \frac{\partial J(\mathbf{a},\mathbf{b})}{\partial a_1} & \cdots & \frac{\partial J(\mathbf{a},\mathbf{b})}{\partial a_d} \end{bmatrix}^\top,$$

where

$$\frac{\partial J(\mathbf{a}, \mathbf{b})}{\partial a_j} = \sum_{i=1}^{n} \frac{\partial (y_i - \hat{y}_i)^2}{\partial a_j} = -2 \sum_{i=1}^{n} \partial (y_i - \hat{y}_i) \frac{\partial \hat{y}_i}{\partial a_j} = 2 \sum_{i=1}^{n} \partial (y_i - \hat{y}_i) b_j x_i e^{-a_j x_i}$$

.