EECS 545 F17 HW2. Name: Po-Jen Lai Members: Ming-Ko Cho, Yi-Lun Wu, Chun-Yu Hsiung

1) (a)

(i) $f(x) = e^x - 1$, $\nabla^2 f(x) = e^x > 0$ for all x.

Also, $f(x)$ is twice continuously differentiable $\Rightarrow$ $f(x)$ is strictly convex and also strictly concave #

(ii)

$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow$ neither PSD or NSD. $\Rightarrow$ neither concex nor concave #

It's strictly concave # because $\{(x_1, x_2) \in \mathbb{R}^2_{++} \mid x_1 x_2 \geq \alpha\}$ are convex. It's not strictly convex #

(iii)

$$\nabla^2 f(x) = \begin{bmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{(1-\alpha)} & \alpha(\alpha-1)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(\alpha-1)x_1^{\alpha-1}x_2^{-\alpha} & (1-\alpha)(-\alpha)x_1^{\alpha}x_2^{-\alpha-1} \end{bmatrix} = -\alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha}\begin{bmatrix} x_1^{-1} \\ -x_2^{-1} \end{bmatrix}\begin{bmatrix} x_1^{-1} \\ -x_2^{-1} \end{bmatrix}^T$$

$\Rightarrow -\nabla^2 f(x) \succeq 0 \Rightarrow f$ is concave.

(b)

(i)

$$J(\theta) = \frac{1}{2}(A^T\theta - y)^T(A^T\theta - y) = \frac{1}{2}\left(\theta^T A A^T \theta - \theta^T A y - y^T A^T \theta - y^T y\right)$$

$\nabla_\theta J(\theta) = AA^T\theta - Ay$ #, $\theta^{(t+1)} = \theta^{(t)} - \alpha(AA^T\theta - Ay)$ #

(ii)

To minimize $J \Rightarrow \nabla_\theta J(\theta) = 0 \Rightarrow AA^T\theta = Ay \Rightarrow \theta = (A^T)^{-1}y$. #

2)

N: numbers of skip.

W: Win.

We know $P(W|N=0) = \frac{1}{2}$, now we assume k Rose Gold have been skipped.

$\Rightarrow P(W|N=n) = \dfrac{50-k}{100-n}$.

$$P(W|N=n+1) = \frac{50-k}{100-(n+1)} \cdot \frac{100-n-(50-k)}{100-n} + \frac{49-k}{100-(n+1)} \cdot \frac{50-k}{100-n} = \frac{50-k}{100-n}.$$

$\Rightarrow P(W|N=n) = P(W|N=n+1) = P(W|N=0) = \frac{1}{2}$ #

$\therefore$ The probability of correct prediction is always $\frac{1}{2}$ regardless of N.

3.

- Matlab code

```
%Training
filename_Y = 'DENSE.Y.TRAIN.Y';
filename_X = 'DENSE.TRAIN.X';
Y = importdata(filename_Y);
X = importdata(filename_X);
count = zeros(2,size(X,2));
count(1,:) = sum(X(Y == 1,:));
count(2,:) = sum(X(Y == -1,:));
p_w_y = zeros(size(count,2),size(count,1));
p_w_y(:,1) = (1 + count(1,:))/(size(count,2) + sum(count(1,:)));
p_w_y(:,2) = (1 + count(2,:))/(size(count,2) + sum(count(2,:)));
py = [sum(Y == 1)/size(Y,1), sum(Y == -1)/size(Y,1)];
%Testing
filename_Ytest = 'DENSE.TEST.Y';
filename_Xtest = 'DENSE.TEST.X';
Y_test = importdata(filename_Ytest);
X_test = importdata(filename_Xtest);
Test_result = zeros(size(Y_test));
prob = zeros(size(Y_test,1),2);
prob(:,1) = log10(py(1)) + X_test*log10(p_w_y(:,1));
prob(:,2) = log10(py(2)) + X_test*log10(p_w_y(:,2));
Test_result = 1*(prob(:,1) > prob(:,2));
Test_result(Test_result == 0) = -1;
error = sum(Test_result ~= Y_test)/size(Y_test,1);
token_ratio = log10(p_w_y(:,1))- log10(p_w_y(:,2));
[n,I] = sort(token_ratio,'descend');
token = importdata('TOKENS_LIST');
Indicator = token(I(1:5));
disp(Indicator);
```

(a) Error rate

Size of training set	50	100	200	400	800	1400
Err rate	0.0388	0.0263	0.0263	0.0188	0.0175	00163

(b) Five tokens : 'httpaddr', 'spam', 'unsubscrib', 'ebai', ' valet'.

4.

- Matlab code

```
load mnist_data.mat;
K = [1 5 9 13];
test_sample_index = randsample(size(test,1),100);
class_l2 = zeros(100,4);
class_l1 = zeros(100,4);
for i = 1:100
    index = test_sample_index(i);
    diff = train(:,2:end) - repmat(test(index,2:end),size(train,1),1);
    diff_l2 = sqrt(sum(diff.^2,2)); % L2-norm
    diff_l1 = sum(abs(diff),2); % L1-norm
    [B2 I2] = sort(diff_l2);
    [B1 I1] = sort(diff_l1);
    for j = 1:size(K,2)
        class_l2(i,j) = mode(train(I2(1:K(j)),1));
        class_l1(i,j) = mode(train(I1(1:K(j)),1));
    end
end
accuracy_rate_l2 = zeros(1,4);
accuracy_rate_l1 = zeros(1,4);
for n = 1:4
    accuracy_rate_l2(n) = sum(class_l2(:,n) == test(test_sample_index,1))/100;
    accuracy_rate_l1(n) = sum(class_11(:,n) == test(test_sample_index,1))/100;
end
```

(c) For K = 1, 5, 9, 13, accuracy rate = 0.95, 0.95, 0.94, 0.94. The K value with best performance are 1,9. However, the result varies while choosing different test samples.

(d) For K = 1, 5, 9, 13, accuracy rate = 0.93, 0.95, 0.94, 0.94. The K value with best performance are 9. However, the result varies when we choose different test samples.

In this problem, I'll choose L2-norm because it always outperforms the L1-norm in terms of accuracy rate.