1、

1. 记正例预测值大于反例则为加一，相等则加 $\frac{1}{2}$，

则对 D^+ 和 D^- 进行遍历则为 $\sum_{x^+\in D^+}\sum_{x^-\in D^-}\left(\mathbb{I}[f(x^+)>f(x^-)]+\frac{1}{2}\mathbb{I}[f(x^+)=f(x^-)]\right)$

又因为此处记得的"1"和"$\frac{1}{2}$"对应于图像中的1单元和$\frac{1}{2}$单元.

且 1单元 $=\frac{1}{m^+m^-}$

$\therefore AUC=\frac{1}{m^+m^-}\sum_{x^+\in D^+}\sum_{x^-\in D^-}\left\{\mathbb{I}[f(x^+)>f(x^-)]+\frac{1}{2}\mathbb{I}[f(x^+)=f(x^-)]\right\}$

2、选做题:

在闭式解方法和牛顿法下，当阈值增大时，预测结果为 0 的个数会增加，预测结果为 1 的个数会减少；反之，则预测结果为 0 的个数会减少，预测结果为 1 的个数会增加。因此，分类阈值的改变直接影响了准确率，查准率和查全率。虽然通常情况下，都会使用 0.5 作为分类阈值，但根据本次实验的实际情况，使用闭式解训练所得分类器在 0.52 附近取得了比 0.5 更好的分类效果，而使用牛顿法训练所得分类器在 0.5 处已取得了很好的分类效果。

3、

```python
'''
Friedman检验不能判断算法性能相同
根据Nemenyi后续检得到：C算法和D算法有显著差别
代码如下：
'''
import numpy as np

data = np.array([[2, 3, 1, 5, 4], [5, 4, 2, 3, 1], [4, 5, 1, 2, 3], [2, 3, 1, 5, 4], [3, 4, 1, 5, 2]])
# Friedman检验
def Friedman(n, k, matrix): # n为数据集个数，k为算法个数，matrix为表格
    # 计算每个算法的平均序值
    row, col = matrix.shape
    value_mean = list()
    for i in range(col):
        value_mean.append(matrix[:, i].mean())

    sum_mean = np.array(value_mean)

    # 计算总的排序和即西伽马ri^2
    sum_ri2_mean = (sum_mean ** 2).sum()
    # 计算Tf
    result_Tx2 = (12 * n) * (sum_ri2_mean - ((k * (k + 1) ** 2) / 4)) / (k * (k + 1))
    result_Tf = (n - 1) * result_Tx2 / (n * (k - 1) - result_Tx2)

    return result_Tf

Tf = Friedman(5, 5, data)
print("Tf = %s" % Tf)
# 输出结果：Tf = 3.9365079365079363

# Nemenyi后续检验
def Nemenyi(n, k, q): # n为数据集个数，k为算法个数，q为检验常用的qa值
    result = q * (np.sqrt(k * (k + 1) / (6 * n)))
    return result
CD = Nemenyi(5, 5, 2.728)
print("CD = %s" % CD)
# 输出为：CD = 2.728
```

4、

4. 假定学习率为η，参数更新估计式为 $v \leftarrow v + \Delta v$.

∴对于 w_{hj} 有：

$$\Delta w_{hj} = -\eta \cdot \frac{\partial E_k}{\partial w_{hj}}$$

又 $E_k = \frac{1}{2}\sum_{j=1}^{l}(\hat{y}_j^k - y_j^k)^2$，$\hat{y}_j^k = f(\beta_j - \theta_j)$，$\beta_j = \sum_{h=1}^{q} w_{hj} b_h$.

$$\therefore \Delta w_{hj} = -\eta \cdot \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial w_{hj}}$$

∵ $\frac{\partial \beta_j}{\partial w_{hj}} = b_h$，sigmoid函数有 $f'(x) = f(x)[1-f(x)]$

$$\therefore \Delta w_{hj} = -\eta \cdot b_h \cdot (\hat{y}_j^k - y_j^k) \cdot f'(\beta_j - \theta_j)$$

$$= -\eta \cdot b_h \cdot (\hat{y}_j^k - y_j^k) \cdot \hat{y}_j^k \cdot (1-\hat{y}_j^k)$$

令 $g_j = -(\hat{y}_j^k - y_j^k)\cdot(1-\hat{y}_j^k)\cdot\hat{y}_j^k = (y_j^k - \hat{y}_j^k)\cdot(1-\hat{y}_j^k)\cdot\hat{y}_j^k$

则 $\Delta w_{hj} = \eta g_j b_h$.

对于 θ_j：

$$\Delta\theta_j = -\eta \cdot \frac{\partial E_k}{\partial \theta_j} = -\eta \cdot \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \theta_j} =$$

同 Δw_{hj} 理，仅对 $\hat{y}_j^k$ 求偏导时需有对"$-\theta$"项求导所得的负

$$\therefore \Delta\theta_j = \eta \cdot g_j \cdot (-1) = -\eta g_j$$

对 v_{ih}：

$$\Delta v_{ih} = -\eta \cdot \frac{\partial E_k}{\partial v_{ih}} = -\eta\left(\sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h}\right)\frac{\partial b_h}{\partial v_{ih}}.$$

$$= -\eta \cdot \left[\sum_{j=1}^{l}\left(\frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j}\right) \cdot w_{hj}\right] \cdot \frac{\partial f(\alpha_h - \gamma_h)}{\partial \alpha_h} \cdot \frac{\partial \alpha_h}{\partial v_{ih}}$$

$$= \eta \cdot \left(\sum_{j=1}^{l} g_j \cdot w_{hj}\right) \cdot b_h \cdot (1-b_h) \cdot x_i$$

∴令 $e_h = \left(\sum_{j=1}^{l} g_j \cdot w_{hj}\right) \cdot b_h \cdot (1-b_h)$

则 $\Delta v_{ih} = \eta \cdot e_h \cdot x_i$

对 γ_h：

$$\Delta\gamma_h = -\eta \cdot \frac{\partial E_k}{\partial \gamma_h} = -\eta\left(\sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h}\right) \cdot \frac{\partial b_h}{\partial \gamma_h}.$$

$$= \eta \cdot \left(\sum_{j=1}^{l} g_j \cdot w_{hj}\right) \cdot \frac{\partial f(\alpha_h - \gamma_h)}{\partial \gamma_h}$$

$$= \eta \cdot \left(\sum_{j=1}^{l} g_j w_{hj}\right) \cdot b_h \cdot (1-b_h) \cdot (-1).$$

$$= -\eta \cdot e_h.$$