Homework #3

---

1. 46
   将数据带入公式计算就行了。
2. [a][d][e]
   假定 $N > d+1$。
   X^TX 可逆 $\Leftrightarrow X$ 各列线性无关。
   $H = X(X^TX)^{-1}X^T$ 是投影矩阵，H 将N维向量投影到 X 的列空间中，$HX = X$。
   H 是实对称矩阵 $\Rightarrow H$ 可对角化 $\Leftrightarrow H$ 特征值的重数等于该特征值的线性无关的特征向量的个数。
   综合上述定理，可以推得：
   H 有 $d+1$ 个特征值为1。[d]正确
   $HH = H$。[e]正确
   H 是实对称矩阵 $\Rightarrow H$ 有N个正交的特征向量。其中，$d+1$ 个特征向量的特征值是1，它们都位于 X 的列空间中，并可以生成 X 的列空间。那么，其余的 $N-(d+1)$ 个特征向量与 X 的列空间正交，所以它们都是 H 特征值为0的特征向量。
   综上，H 有 $d+1$ 个特征值为1，$N-(d+1)$ 个特征值为0。[a]正确
3. [a][b][e]
   仿照Lecture11课件中那样画图，就能看出来了。
4. [b][d][e]
5. 证明如下：

$$err(w) = \max(0, -yw^Tx) = \begin{cases} -yw^Tx, & yw^Tx < 0 \\ 0, & yw^Tx \geq 0 \end{cases}$$

$$\nabla_{w_t} err(w_t, x_n, y_n) = \begin{cases} -y_n x_n, & y_n w^T x_n < 0 \\ 0, & y_n w^T x_n \geq 0. \end{cases}$$

$$w_{t+1} = w_t + \eta[-\nabla_{w_t}(w_t, x_n, y_n)] = \begin{cases} w_t + \eta y_n x_n, & y_n \neq sign(w^T x_n) \\ w_t, & y_n = sign(w^T x_n) \text{ 或 } w^T x_n = 0. \end{cases}$$

6. $(-2, 0)$
7. 2.825
8. $(1.5, 4, -1, -2, 0, 3)$
9. 计算过程如下:

令

$$\begin{cases} \dfrac{\partial \hat{E}_2(\Delta u,\Delta v)}{\partial(\Delta u)} = 0 = E_{uu}(u,v)\cdot(\Delta u) + E_{uv}(u,v)\cdot(\Delta v) + E_u(u,v) \\ \dfrac{\partial \hat{E}_2(\Delta u,\Delta v)}{\partial(\Delta v)} = 0 = E_{vv}(u,v)\cdot(\Delta v) + E_{uv}(u,v)\cdot(\Delta u) + E_v(u,v). \end{cases}$$

记 $A_1 = E_{uu}(u,v)$, $A_2 = E_{uv}(u,v)$, $E_u(u,v) = A_3$,

$B_1 = E_{vv}(u,v)$, $B_3 = E_v(u,v)$.

$x = \Delta u$, $y = \Delta v$.

上式可表示为 $\begin{cases} A_1 x + A_2 y + A_3 = 0 \\ B_1 y + A_2 x + B_3 = 0, \end{cases}$ 即 $\begin{bmatrix} A_1 & A_2 \\ A_2 & B_1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} A_3 \\ B_3 \end{bmatrix}$.

由题可知 Hessian matrix 是正定矩阵，所以 $\begin{bmatrix} A_1 & A_2 \\ A_2 & B_1 \end{bmatrix}$ 可逆

$$\begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} A_1 & A_2 \\ A_2 & B_1 \end{bmatrix}^{-1}\begin{bmatrix} A_3 \\ B_3 \end{bmatrix} = -(\nabla^2 E(u,v))^{-1}\nabla E(u,v).$$

根据多元函数的极值相关定理，可知 $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = -(\nabla^2 E(u,v))^{-1}\nabla E(u,v)$ 是最小值点。

- 参考资料：
  - Jacobian矩阵和Hessian矩阵
  - 使用牛顿法寻找极值

10. 2.362
    (五次迭代后，用牛顿法得到的值更小，说明牛顿法比梯度下降法速度更快。)
11. X1 X2 X3 X4 X5 X6
    假设 $x = (x_1, x_2)$，特征转换 $\Phi_2(x) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2) = z$。
    可求得：

$$\begin{aligned}
\Phi_2(x_1) &= (1,1,1,1,1,1) && = z_1 \\
\Phi_2(x_2) &= (1,1,-1,1,-1,1) && = z_2 \\
\Phi_2(x_3) &= (1,-1,-1,1,1,1) && = z_3 \\
\Phi_2(x_4) &= (1,-1,1,1,-1,1) && = z_4 \\
\Phi_2(x_5) &= (1,0,0,0,0,0) && = z_5 \\
\Phi_2(x_6) &= (1,1,0,1,0,0) && = z_6
\end{aligned}$$

令

$$Z=\begin{bmatrix} z_1^T \\ z_2^T \\ z_3^T \\ z_4^T \\ z_5^T \\ z_6^T \end{bmatrix}=\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}, W=\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix}$$

可求得 Z可逆。

令 $ZW=y$,对于任意 $y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_6 \end{bmatrix}$，都有$W=Z^{-1}y$。那么，$z_1,z_2,z_3,z_4,z_5,z_6$ 可以被shatter，所以 x_1,x_2,x_3,x_4,x_5,x_6 也可以被shatter。

12. ∞
题中特征转换的意思是，对于data中的任一 x，

$$\Phi(x)=z=\begin{bmatrix} [x=x_1] \\ [x=x_2] \\ \vdots \\ [x=x_N] \end{bmatrix}$$

其中，$[x=x_n], n=1,2,\cdots,N$ 的意思是：若$x=x_n$成立，则值为1，否则为0。

若$x_i(i=1,2,\cdots,N)$都不相同，那么，$\Phi(x_1)=z_1=\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$，

$\Phi(x_2)=z_2=\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$，$\cdots$，$\Phi(x_N)=z_N=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$。

令$Z=\begin{bmatrix} 1 & z_1^T \\ 1 & z_2^T \\ \vdots & \vdots \\ 1 & z_N^T \end{bmatrix}$，可知 Z的秩为N，行满秩。令 $ZW=y$，对于任意$y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$，$W$都有解。也就是说，$z_1,z_2,\cdots,z_N$ 可以被shatter。所以，d_{vc} 是∞。

- 13-15：编程题

16. 和page 11 of Lecture 10 slides的推导过程差不多，相乘、取对数、加负号。

$$\text{likelihood}((w_1, \cdots, w_K)) \propto \prod_{n=1}^{N} h_{y_n}(x_n) = \prod_{n=1}^{N} \frac{\exp(W_{y_n}^T x_n)}{\sum_{i=1}^{K} \exp(W_i^T x_n)}$$

$$\begin{aligned} E_{in}((w_1, \cdots, W_K)) &= -\frac{1}{N} \ln \prod_{n=1}^{N} \frac{\exp(W_{y_n}^T x_n)}{\sum_{i=1}^{K} \exp(W_i^T x_n)} \\ &= \frac{1}{N} \sum_{n=1}^{N} \ln \frac{\sum_{i=1}^{K} \exp(W_i^T x_n)}{\exp(W_{y_n}^T x_n)} \\ &= \frac{1}{N} \sum_{n=1}^{N} \left[\ln \left(\sum_{i=1}^{K} \exp(W_i^T x_n) \right) - W_{y_n}^T x_n \right]. \end{aligned}$$

17.

$$\begin{aligned} \frac{\partial E_{in}}{\partial w_i} &= \frac{1}{N} \sum_{n=1}^{N} \left(\frac{exp(w_i^T x_n)}{\sum_{i=1}^{K} exp(w_i^T x_n)} x_n - [y_n = i] x_n \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} ((h_i(x_n) - [y_n = i]) x_n) \end{aligned}$$

- **18-20**：编程题