Homework #4

1. In general,deterministic noise will increase.
   Deterministic noise: difference between best $h^* \in H$ and f,if $f \notin H$.
2. H_2
   $H(10,0,3) \cap H(10,0,4) = H(10,0,3) = H_2$
3. $\alpha = 1 - \frac{2\lambda\eta}{N}, \beta = -\eta$

$$
\begin{aligned}
W_{t+1} &= W_t - \eta \nabla E_{aug}(w_t) \\
\nabla E_{aug}(w) &= \nabla E_{in}(w) + \nabla(\frac{\lambda}{N} W^T W) \\
&= \nabla E_{in}(w) + \frac{2\lambda}{N} W \\
W_{t+1} &= (1 - \frac{2\lambda\eta}{N}) W_t - \eta \nabla E_{in}(w_t)
\end{aligned}
$$

4.

$W_{reg}(\lambda) = \underset{W}{\operatorname{argmin}} E_{aug}(W) = (Z^T Z + \lambda I)^{-1} Z^T y$, for $\lambda > 0$.

$\|W_{reg}(\lambda)\| = [W_{reg}(\lambda)]^T [W_{reg}(\lambda)] = [(Z^T Z + \lambda I)^{-1} Z^T y]^T [(Z^T Z + \lambda I)^{-1} \cdot Z^T y]$

$Z^T Z$ 是实对称矩阵，且为半正定矩阵，那么

$$
Z^T Z = Q \Lambda Q^{-1} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & q_{d+1} \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{d+1} \end{bmatrix} \begin{bmatrix} \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & q_{d+1} \\ \vdots & \vdots & & \vdots \end{bmatrix}^{-1},
$$

其中，$q_i (1 \le i \le d+1)$ 是 $Z^T Z$ 的特征向量，它的特征值为 λ_i。并且，

$$
q_i^T q_j = \begin{cases} 1, \text{if } i = j \\ 0, \text{if } i \ne j \end{cases}, \quad \lambda_i \ge 0 \ (1 \le i \le d+1)
$$

当$\lambda>0$，$Z^TZ+\lambda I$ 是实对称矩阵，正定矩阵，~~那么~~可作如下分解：

$$Z^TZ+\lambda I = Q\Lambda_1 Q^{-1},\quad Q\text{同上},\quad \Lambda_1=\begin{bmatrix}\lambda_1+\lambda & & & \\ & \lambda_2+\lambda & & \\ & & \ddots & \\ & & & \lambda_{d+1}+\lambda\end{bmatrix}$$

$Z^TZ+\lambda I$ 的逆矩阵 $(Z^TZ+\lambda I)^{-1}$ 也是正定矩阵，可作分解：

$$(Z^TZ+\lambda I)^{-1} = Q\Lambda_2 Q^{-1},\quad Q\text{同上},\quad \Lambda_2=\begin{bmatrix}\frac{1}{\lambda_1+\lambda} & & & \\ & \frac{1}{\lambda_2+\lambda} & & \\ & & \ddots & \\ & & & \frac{1}{\lambda_{d+1}+\lambda}\end{bmatrix}$$

由于Q的各个列向量相互正交，那么 Z^Ty 可由Q的列向量线性表出：

$$Z^Ty = Q\begin{bmatrix}k_1\\ \vdots\\ k_{d+1}\end{bmatrix} = \begin{bmatrix}\vdots & & \vdots\\ q_1 & \cdots & q_{d+1}\\ \vdots & & \vdots\end{bmatrix}\begin{bmatrix}k_1\\ \vdots\\ k_{d+1}\end{bmatrix}$$

$$(Z^TZ+\lambda I)^{-1}(Z^Ty) = (Z^TZ+\lambda I)^{-1}\cdot Q\begin{bmatrix}k_1\\ \vdots\\ k_{d+1}\end{bmatrix} = Q\Lambda_2\begin{bmatrix}k_1\\ \vdots\\ k_{d+1}\end{bmatrix} = Q\begin{bmatrix}k_1/\lambda_1+\lambda\\ \vdots\\ k_{d+1}/\lambda_{d+1}+\lambda\end{bmatrix}$$

$$\left[(Z^TZ+\lambda I)^{-1}(Z^Ty)\right]^T = \begin{bmatrix}\frac{k_1}{\lambda_1+\lambda} & \cdots & \frac{k_{d+1}}{\lambda_{d+1}+\lambda}\end{bmatrix}Q^T.$$

$$\begin{aligned}\|W_{reg}(\lambda)\| &= \left[(Z^TZ+\lambda I)^{-1}Z^Ty\right]^T\left[(Z^TZ+\lambda I)^{-1}Z^Ty\right]\\ &= \begin{bmatrix}\frac{k_1}{\lambda_1+\lambda} & \cdots & \frac{k_{d+1}}{\lambda_{d+1}+\lambda}\end{bmatrix}Q^TQ\begin{bmatrix}k_1/(\lambda_1+\lambda)\\ \vdots\\ k_{d+1}/(\lambda_{d+1}+\lambda)\end{bmatrix}\\ &= \frac{k_1^2}{(\lambda_1+\lambda)^2}+\cdots+\frac{k_{d+1}^2}{(\lambda_{d+1}+\lambda)^2}.\end{aligned}$$

5. $\sqrt{9+4\sqrt{6}}$
   - 设$x_1=(-1,0),x_2=(\rho,1),x_3=(1,0)$
   - **model 1**: $[h_0(x)=b_0]$
     $g_1^-(x)=0.5,e_1=0.25;g_2^-(x)=0,e_2=1;g_3^-(x)=0.5,e_3=0.25$
     $(e_1+e_2+e_3)/3=0.5$
   - **model 2**:$[h_1(x)=a_1x+b_1]$
     $g_1^-(x)=\frac{1}{\rho-1}x-\frac{1}{\rho-1},e_1=\frac{4}{(\rho-1)^2};$
     $g_2^-(x)=0,e_2=1;$
     $g_3^-(x)=\frac{1}{\rho+1}x+\frac{1}{\rho+1},e_3=\frac{4}{(\rho+1)^2}$
     $(e_1+e_2+e_3)/3=[\frac{4}{(\rho-1)^2}+1+\frac{4}{(\rho+1)^2}]/3$
   - 解$0.5=[\frac{4}{(\rho-1)^2}+1+\frac{4}{(\rho+1)^2}]/3$，得$\rho=\sqrt{9+4\sqrt{6}}$
6. 63
   $32+16+8+4+2+1=63$
7. 370
   $1000-63\times 10=370$

8. 1
   Before you look at the data, you do mathematical derivations and come up with **a credit approval function.** You now **test it on the data** and, to your delight, **obtain perfect prediction.**
9. 0.271
   令N=10000，ϵ=0.01,带入公式$P[|\nu-\mu|>\epsilon] \le 2exp(-2\epsilon^2 N)$计算。
10. a(x) AND g(x)
    同时使用a(x)和g(x)判断，两者都通过才能得到信用卡。因为训练得到g(x)所用的样本是经过a(x)筛选过的，g(x)能在筛选后的数据中做出好的预测。
11. $w=(X^TX+\tilde{X}^T\tilde{X})^{-1}(X^Ty+\tilde{X}^T\tilde{y})$
    按照lecture9 slide第9页的方法求出梯度，然后令梯度为0，求之得optimal w。
12.

$$
\begin{aligned}
&\tilde{X}^T=\sqrt{\lambda}I,\tilde{y}=0 \\
\Rightarrow&\tilde{X}^T\tilde{X}=\lambda I,\tilde{X}^T\tilde{y}=0 \\
\Rightarrow&(X^TX+\tilde{X}^T\tilde{X})^{-1}(X^Ty+\tilde{X}^T\tilde{y})=(X^TX+\lambda I)^{-1}X^Ty
\end{aligned}
$$

- 13至20题：编程题