1，假设训练样本集为 $D=\{(\vec{x}_1,y_1)=((0.2,0.7)^T,1),(\vec{x}_2,y_2)=((0.3,0.3)^T,1),(\vec{x}_3,y_3)=((0.4,0.5)^T,1),(\vec{x}_4,y_4)=((0.6,0.5)^T,1),(\vec{x}_5,y_5)=((0.1,0.4)^T,1),(\vec{x}_6,y_6)=((0.4,0.6)^T,-1),(\vec{x}_7,y_7)=((0.6,0.2)^T,-1),(\vec{x}_8,y_8)=((0.7,0.4)^T,-1),(\vec{x}_9,y_9)=((0.8,0.6)^T,-1),(\vec{x}_{10},y_{10})=((0.7,0.5)^T,-1)\}$，使用线性回归算法（Linear Regression Algorithm），通过广义逆来求解，并设计这两类的分类函数，讨论结果。（可通过编程计算得到广义逆的结果）

2，根据向量或矩阵的计算性质，证明：

$$\|\mathbf{X}\boldsymbol{w}-Y\|^2=\boldsymbol{w}^T\mathbf{X}^T\mathbf{X}\boldsymbol{w}-2\boldsymbol{w}^T\mathbf{X}^T\mathrm{Y}+\mathrm{Y}^T\mathrm{Y}$$

3，总结梯度下降法、随机梯度下降法、Adagrad、RMSProp、动量法（Momentum）和 Adam 等方法权系数更新表达式。

3、① 梯度下降

a. $\nabla L_{in}(w)=\sum_{n=1}^{N}(w^T x_n-y_n)x_n$

b. $w_{t+1}\leftarrow w_t-\eta\nabla L_{in}(w_t)$

② 随机梯度下降

a. $\nabla L_{in}(w)=(w^T x_n-y_n)x_n$

b. $w_{t+1}\leftarrow w_t-\eta\nabla L_{in}(w_t)$

③ Adagrad

a. $\nabla L_{in}(w)=\sum_{n=1}^{B}(w^T x_n-y_n)x_n$

b. $\sigma_{i,t}=\sqrt{\frac{1}{t+1}\sum_{t=0}^{t}\left(\frac{\partial L_{in}}{\partial w_{i,t}}\right)^2}$

c. $w_{i,t+1}\leftarrow w_{i,t+1}-\frac{\eta}{\sigma_{i,t}}\frac{\partial L_{in}}{\partial w_{i,t}}$

④ RMSProp

a. $\nabla L_{in}(w)=\sum_{n=1}^{B}(w^T x_n-y_n)x_n$

b. $\sigma_{i,t}=\sqrt{\alpha(\sigma_{i,t-1})^2+(1-\alpha)\left(\frac{\partial L_{in}}{\partial w_{i,t}}\right)^2}$

c. $w_{i,t+1}=w_{i,t}-\frac{\eta}{\sigma_{i,t}}\frac{\partial L_{in}}{\partial w_{i,t}}$

⑤ Momentum

a. $\nabla L_{in}(w)=\sum_{n=1}^{B}(w^T x_n-y_n)x_n$

b. $m_t=\lambda m_{t-1}-\eta\nabla L_{in}(w_t)$

c. $w_{t+1}\leftarrow w_t+m_t$

⑥ Adam

a. $\nabla L_{in}(w)=\sum_{n=1}^{B}(w^T x_n-y_n)x_n$

b. $m_t=\beta_1 m_{t-1}+(1-\beta_1)\nabla L_{in}(w)$

c. $v_t=\beta_2 v_{t-1}+(1-\beta_1)(\nabla L_{in}(w))^2$

d. $m_t'=\frac{m_t}{1-\beta_1^t}$，$v_t'=\frac{v_t}{1-\beta_2^t}$

e. $w_t\leftarrow w_{t-1}-\alpha\frac{m_t'}{\sqrt{v_t'}+\epsilon}$

1、解：$\because X=\begin{bmatrix}0.2&0.3&0.4&0.6&0.1&0.4&0.6&0.7&0.8&0.7\\0.7&0.3&0.5&0.5&0.4&0.6&0.2&0.4&0.6&0.5\\1&1&1&1&1&1&1&1&1&1\end{bmatrix}^T$

$y=[1\ 1\ 1\ 1\ 1\ -1\ -1\ -1\ -1\ -1]^T$

$\therefore X^+=(X^TX)^{-1}X^T=\begin{bmatrix}5.1&2.42&2.92\\4.7&2.41&2.24\\10&4.7&4.8\end{bmatrix}^{-1}X^T$

$=\begin{bmatrix}-0.07&0.71&0.15&-0.06&0.70&-0.40&0.60&0.05&-0.50&-0.17\\1.09&-0.88&0.13&0.16&-0.41&-0.68&-1.32&-0.31&0.69&0.18\\-0.71&-0.40&-0.24&0.18&-0.85&0.37&0.26&0.42&0.58&0.39\end{bmatrix}$

$\therefore W^*=X^+Y=\begin{bmatrix}1.85\\0.18\\-4.03\end{bmatrix}$　$y_{11}=\begin{cases}1, & -4.03x_1+0.18x_2+1.85>0\\-1, & -4.03x_1+0.18x_2+1.85\leq 0\end{cases}$

上述分类器无法正确分类 $\{(0.6,0.5),1\}$ 与 $\{(0.4,0.6),-1\}$ 两个样本，正确率80%

2、$\|Xw-Y\|^2=(Xw-Y)^T(Xw-Y)=(w^TX^T-Y^T)(Xw-Y)=w^TX^TXw-Y^TXw-w^TX^TY+Y^TY$

$=w^TX^TXw-(w^TX^TY)^T-w^TX^TY+Y^TY$

$\because X:N\times(d+1)$，$w:(d+1)\times 1$，$Y:N\times 1$　$\therefore w^TX^TY$ 为一标量 $\therefore w^TX^TY=(w^TX^TY)^T$

$\therefore \|Xw-Y\|^2=w^TX^TXw-2w^TX^TY+Y^TY$