hw0_sol   Haoyuan Du

Q1.

a.

$$\varphi(\mathrm{w}) = (\mathrm{Xw})^{\mathrm{T}}(\mathrm{Xw}) - 2\mathrm{Xw}^{\mathrm{T}}\mathrm{y} + \mathrm{y}^{\mathrm{T}}\mathrm{y}$$

$$= \mathrm{w}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{Xw} - 2\mathrm{w}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}} + \mathrm{y}^{\mathrm{T}}\mathrm{y}$$

$$\frac{\partial\varphi(\mathrm{w})}{\partial\mathrm{w}} = 2\mathrm{X}^{\mathrm{T}}\mathrm{Xw} - 2\mathrm{X}^{\mathrm{T}}\mathrm{y} = 0$$

$$\mathrm{X}^{\mathrm{T}}\mathrm{Xw} = \mathrm{X}^{\mathrm{T}}\mathrm{y}$$

$$(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}(\mathrm{X}^{\mathrm{T}}\mathrm{X})\mathrm{w} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

$$\mathrm{w} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

Because in equation $\varphi(w) + 2\,[X\delta]^T\,[Xw - y] + (X\delta)^TX\delta$, $[X\delta]^T$ is a 1*n matrix, $[Xw - y]$ is a n*1 matrix, so the product is a 1*1 matrix which is a positive number. Similarly, $(X\delta)^T$ is a 1*n matrix. So $(X\delta)^TX\delta$ is a 1*1 matrix which is a positive number (sum of squares). $\varphi(w)$ plus a positive number will greater than $\varphi(w)$ itself which can't be the optimal solution (minimum). Thus, w must be determined so that $\varphi(w + \delta) \geqslant \varphi(w)$ for any possible vector δ.

b.

$$\frac{\partial\varphi(\mathrm{w})}{\partial\mathrm{w}} = 2\mathrm{X}^{\mathrm{T}}\mathrm{Xw} - 2\mathrm{X}^{\mathrm{T}}\mathrm{y} + 2\lambda\mathrm{Iw} = 0$$

$$\mathrm{X}^{\mathrm{T}}\mathrm{Xw} + \lambda\mathrm{Iw} = \mathrm{X}^{\mathrm{T}}\mathrm{y}$$

Sol:

$$\mathrm{w} = (\mathrm{X}^{\mathrm{T}}\mathrm{X} + \lambda\mathrm{I})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

Q2.

a.

No. Because the probability can't be used to determine a single (individual) event. It can only describe a large group of data.

b.

Table 1

	D	Not D	
Positive+	P (X = 1|D = 1) = 0.7 1000*0.7 = 700	1800	2500
Negative-	300	P (X = 0|D = 0) = 0.8) 9000*0.8 = 7200	7500
	10000*0.1 = 1000	9000	10000

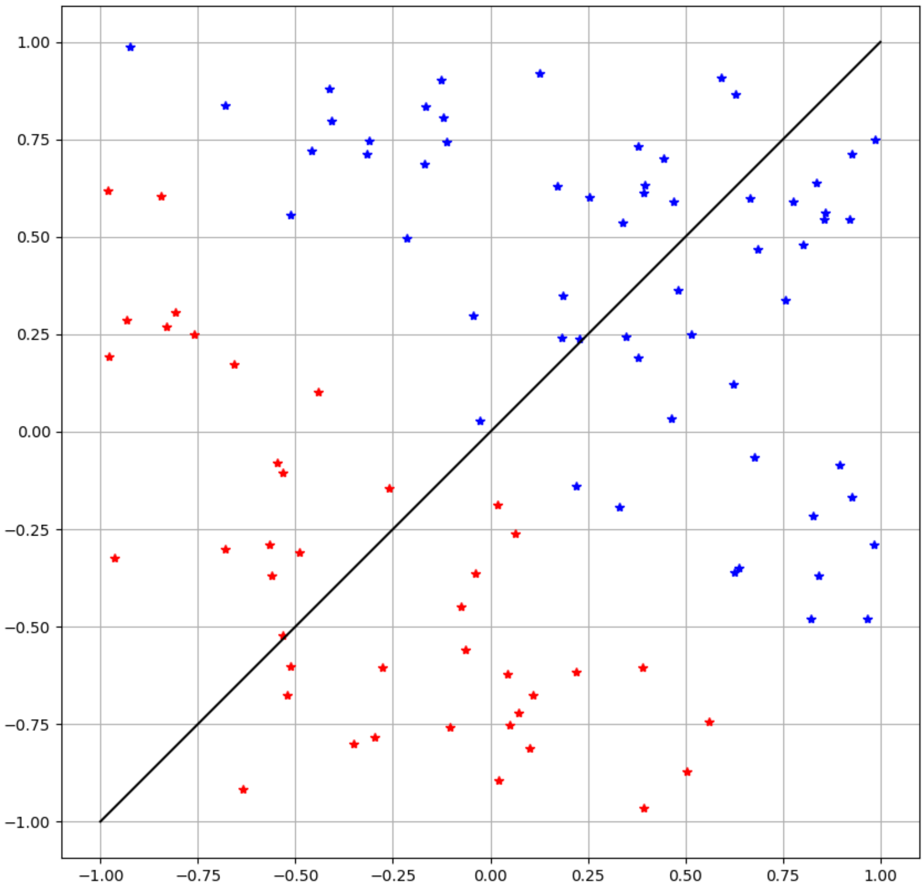
As the table above, P(disease) = P(b) = 0.1, P(positive | disease) = 0.7, P(positive) = P(a) = 2500 / 10000 = 0.25

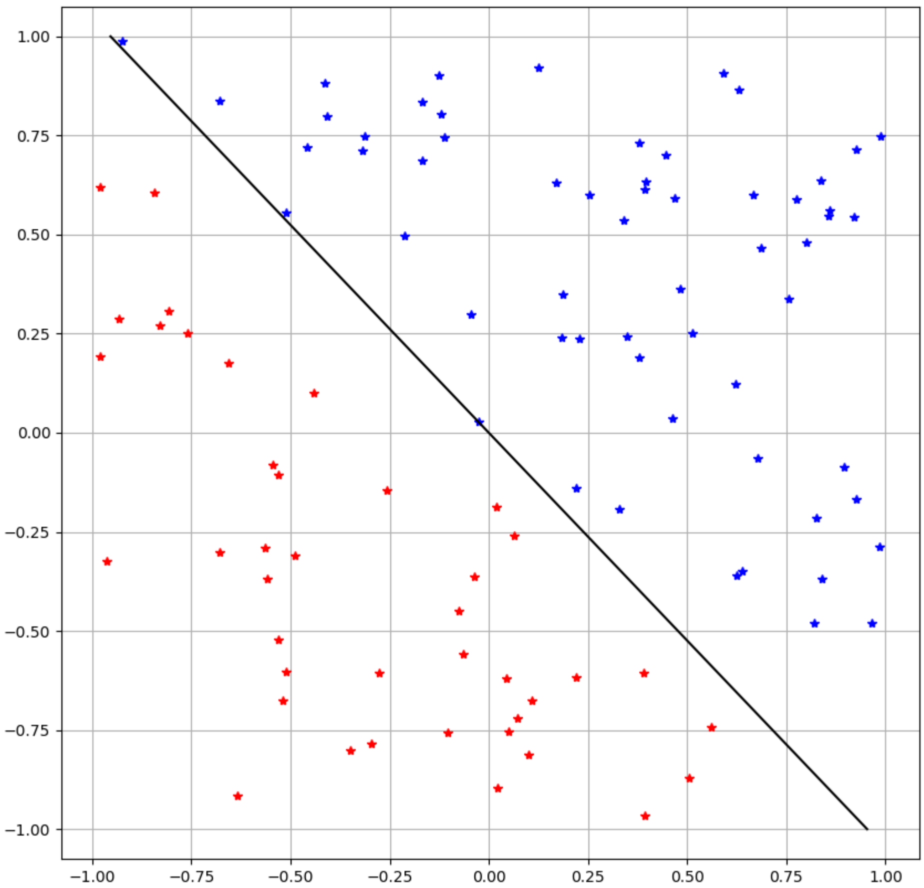
Thus, according to the Bayes Rule: p(b|a) = p(a|b)p(b) / p(a) = 0.7*0.1 / 0.25 = 0.28

We can get the same answer from the table: 700/2500 = 0.28

Q3.

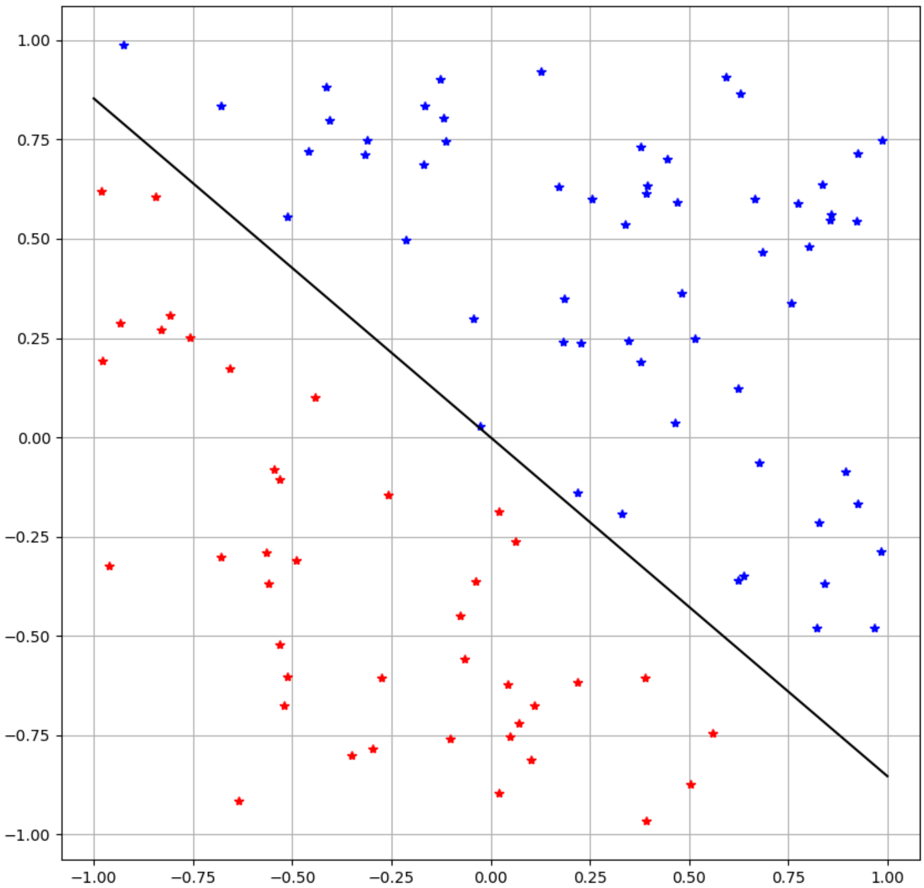
81 iterations for convergence. Error rate = 0.09.





Q4.

error rate = 0.00


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00