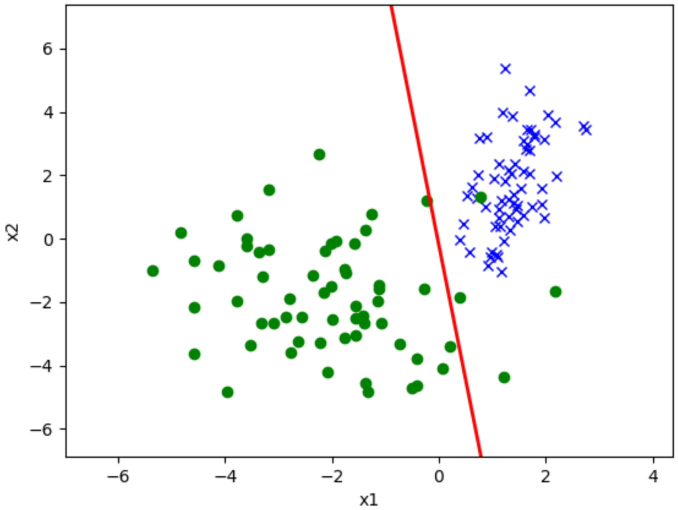


train use y-label, rescale by α

3.

(a)

$$p(y;\lambda) = \frac{1}{y!}\exp\{\log\lambda \cdot y - \lambda\}$$

$$\begin{cases} b(y) & = \frac{1}{y!} \\ \eta & = \log\lambda \\ T(y) & = y \\ a(\eta) & = e^{\eta} \end{cases}$$

(b)

$$h_\theta(x) = E(y|x;\theta) = \lambda = e^{\eta} = e^{\theta^T x}$$

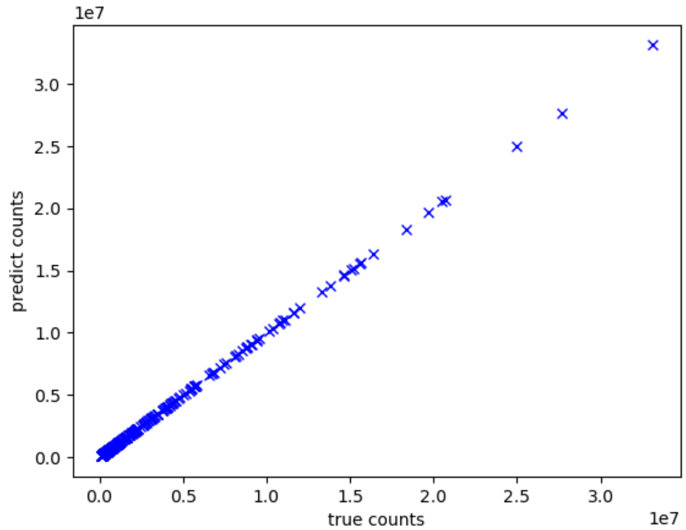
(c)

$$\begin{aligned} \log p(y^{(i)}|x^{(i)};\theta) &= \log\frac{1}{y^{(i)}!}\exp\{\theta^T x^{(i)} y^{(i)} - e^{\theta^T x^{(i)}}\} \\ &= -\log y^{(i)}! + \theta^T x^{(i)} y^{(i)} - e^{\theta^T x^{(i)}} \end{aligned}$$

$$\frac{\partial \log p(y^{(i)}|x^{(i)};\theta)}{\partial\theta_j} = y^{(i)}x_j^{(i)} - e^{\theta^T x^{(i)}} \cdot x_j^{(i)} = (y^{(i)} - e^{\theta^T x^{(i)}})x_j^{(i)}$$

$$\theta_j := \theta_j + \alpha \cdot (y^{(i)} - e^{\theta^T x^{(i)}})x_j^{(i)}$$

(d)


1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
predict counts
0.0
0.5
1.0
1.5
2.0
2.5
3.0
true counts
1e7

4.