1) **Logistic Regression (code in *fisher_scoring.py*)**

a) Find the Hession matrix for

$$J(\theta) = \frac{1}{m}\Sigma_1^m \log\left(1 + e^{-y^{(i)}\theta^T x^{(i)}}\right) = -\frac{1}{m}\Sigma_1^m \log\left(h_\theta\left(y^{(i)}x^{(i)}\right)\right)$$

where $y^{(i)} \in \{-1, 1\}, h_\theta(x) = g(\theta^T x)$ *and* $g(z) = \frac{1}{1+\exp(-z)}$

$$\frac{\partial J}{\partial \theta_i} = -\frac{1}{m}\Sigma_{k=1}^m \left(\frac{1}{h_\theta(y^{(k)}x^{(k)})} \times \frac{\partial h_\theta}{\partial_{\theta_i}}\right) \Rightarrow$$

$$\frac{\partial J}{\partial \theta_i} = -\frac{1}{m}\Sigma_{k=1}^m \left(\frac{g\left(\theta^T y^{(k)}x^{(k)}\right)\left(1 - g\left(\theta^T y^{(k)}x^{(k)}\right)\right)}{h_\theta(y^{(k)}x^{(k)})} \times \frac{\partial\left(\theta^T y^{(k)}x^{(k)}\right)}{\partial_{\theta_i}}\right) \Rightarrow$$

$$\frac{\partial J}{\partial \theta_i} = -\frac{1}{m}\Sigma_{k=1}^m \left(\left(1 - g\left(\theta^T y^{(k)}x^{(k)}\right)\right) y^{(k)}x_i^{(k)}\right) \Rightarrow$$

$$\frac{\partial^2 J}{\partial_{\theta_i}\partial_{\theta_j}} = \frac{1}{m}\Sigma_{k=1}^m \left(g\left(\theta^T y^{(k)}x^{(k)}\right)\left(1 - g\left(\theta^T y^{(k)}x^{(k)}\right)\right) y^{(k)}x_i^{(k)} \times \frac{\partial\left(\theta^T y^{(k)}x^{(k)}\right)}{\partial_{\theta_i}}\right) \Rightarrow$$

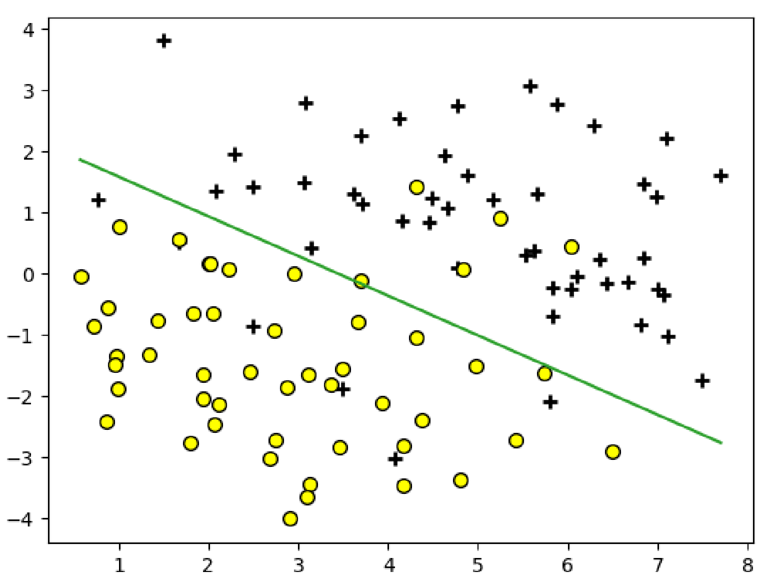
Hessian matrix: $H_{ij} = \frac{1}{m}\Sigma_{k=0}^m \left(g\left(\theta^T y^{(k)}x^{(k)}\right)\left(1 - g\left(\theta^T y^{(k)}x^{(k)}\right)\right) y^{(k)}y^{(k)}x_i^{(k)}x_j^{(k)}\right)$

Show that H is ***semi positive definite***: Suppose $A = diag\left(\left[A^{(1)}\, A^{(n)} \ldots A^{(m)}\right]\right), \boldsymbol{A^{(k)}} * \boldsymbol{A^{(k)}} = g\left(\theta^T y^{(k)}x^{(k)}\right)\left(1 - g\left(\theta^T y^{(k)}x^{(k)}\right)\right) \geq 0, then\ A^{(k)} \geq 0$;

Then $z^T H z = \Sigma_i \Sigma_j z_i\left(x_i A A x_j\right) z_j = (Axz)^2 \geq 0$

b) Fitted coefficients θ = [-2.6205116   0.76037154 1.17194674]

c) The figure plotted in ps1q1c.png



2) Poisson regression and the exponential family

a) Show that Poisson distribution $p(y, \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$ is in exponential family.

$$p(y, \lambda) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{1}{y!}\exp(log(\lambda)y - \lambda)$$
$$\eta = \log(\lambda)$$
$$T(y) = y$$
$$a(\eta) = \lambda$$
$$b(y) = \frac{1}{y!}$$

b) According to the canonical parameter $\eta = \log(\lambda)$, the canonical response function (link function) is its inverse $\lambda = \exp(\eta)$, plug in $\eta = \theta^T x$ the final canonical response function is $\lambda = \exp(\theta^T x)$

c) Plug in the canonical response function into $p(y, \lambda)$ and take the derivative of its log with respect to θ:

$$p(y|x;\theta) = \frac{1}{y!}\exp(y\theta^T x - \exp(\theta^T x))$$
$$l(y|x;\theta) = log\frac{1}{y!} + y\theta^T x - \exp(\theta^T x)$$
$$\frac{\partial l}{\partial \theta_i} = yx_i - x_i\exp(\theta^T x) = x_i(y - \exp(\theta^T x))$$

The stochastic gradient ascent rule is $\theta_i \coloneqq \theta_i + \alpha\left(y - e^{\theta^T x}\right)x_i$

d) Show that for any Exponential Family distribution the max log-likelihood stochastic gradient ascent will result in update rule $\theta_i \coloneqq \theta_i + \alpha\big(y - h_\theta(x)\big)x_i$

Prove:

$$p(y{:}\,\eta) = b(y)\exp\Big(\eta^T T(y) - a(\eta)\Big) \Rightarrow$$
$$l(y{:}\,\eta) = \log\big(b(y)\big) + \eta^T T(y) - a(\eta) \Rightarrow$$
$$\frac{\partial l}{\partial \theta_i} = yx_i - a'(\eta)x_i = \big(y - a'(\eta)\big)x_i$$

Compare it with the update rule the question now is to prove:

$$E[y|x;\theta] = h_\theta(x) = a'(\eta)$$

Use the fact that $\int p(y)\,dy = 1$

$$f(\eta) = \int b(y)\exp(\eta^T y - a(\eta))dy = 1 \Rightarrow$$
$$\frac{df(\eta)}{d\eta} = \int b(y)\exp\big(\eta^T y - a(\eta)\big)\big(y - a'(\eta)\big)dy = 0 \Rightarrow$$

$$\int \big(y - a'(\eta)\big)p(y)\,dy = 0 \Rightarrow$$

$$\int yp(y)\,dy - \int a'(\eta)p(y)\,dy = 0 \Rightarrow$$

$$E[y] = h_\theta(x) = a'(\eta)\int p(y)dy = a'(\eta) \Rightarrow proved.$$

3) Gaussian discriminant analysis

$$p(y) = \begin{cases} \phi & if\ y = 1 \\ 1-\phi & if\ y = -1 \end{cases}$$

$$p(x|y=-1) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_{-1})^T\Sigma^{-1}(x-\mu_{-1})\right)$$

$$p(x|y=1) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_{1})^T\Sigma^{-1}(x-\mu_{1})\right)$$

Show that the posterior distribution of the label at x takes the form of a logistic function, and can be written: $p(y\,|\,x;\phi,\Sigma,\mu\_(-1),\mu\_1\,) = 1/(1+\exp(-y(\theta\hat{}T\ x+\theta_0))\,)$

Try some brute force math, $p(y=0)$ is the same: $p(y=1\,|\,x) = (p(x\,|\,y=1)p(y=1))/p(x)$

$$= \frac{\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)\phi}{\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\left(\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)\phi + \exp\left(-\frac{1}{2}(x-\mu_{-1})^T\Sigma^{-1}(x-\mu_{-1})\right)(1-\phi)\right)}$$

$$= \frac{\phi\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)}{\phi\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right) + (1-\phi)\exp\left(-\frac{1}{2}(x-\mu_{-1})^T\Sigma^{-1}(x-\mu_{-1})\right)}$$

$$= \frac{1}{1+\frac{(1-\phi)}{\phi}\exp\left(\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - \frac{1}{2}(x-\mu_{-1})^T\Sigma^{-1}(x-\mu_{-1})\right)}$$

$$= \frac{1}{1+\frac{(1-\phi)}{\phi}\exp\left(\frac{1}{2}(2u_{-1}\Sigma^{-1}x - 2u_1\Sigma^{-1}x - (u_{-1}^T\Sigma^{-1}u_{-1} - u_1^T\Sigma^{-1}u_1))\right)}$$

$$= \frac{1}{1+\exp\left((u_{-1}-u_1)\Sigma^{-1}x - \frac{1}{2}(u_{-1}^T\Sigma^{-1}u_{-1} - u_1^T\Sigma^{-1}u_1) + \log\left(\frac{1-\phi}{\phi}\right)\right)}$$

Set $\theta = (u_{-1} - \mathrm{u}_1)\Sigma^{-1} and\ \theta_0 = \log\left(\frac{1-\phi}{\phi}\right) - \frac{1}{2}(u_{-1}^T\Sigma^{-1}u_{-1} - u_1^T\Sigma^{-1}u_1) \Rightarrow$

$$p(y = 1|x) = \frac{p(x|y = 1)p(y = 1)}{p(x)} \Rightarrow PROVED$$

4)