CS 229, Fall 2018
Problem Set 1

August 7, 2023

1 Linear Classifiers (logistic regression and GDA)

1.1 a

$$
\begin{aligned}
J(\theta) &= -\frac{1}{m}\sum_{i=1}^{m}(y^i log(g(\theta^T x)) + (1-y^i)log(1-g(\theta^T x))) \\
g'(x) &= g(x)(1-g(x)) \\
\frac{\partial J(\theta)}{\partial \theta_i} &= -\frac{1}{m}(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)})\frac{\partial g(\theta^T x)}{\partial \theta_i} \\
&= -\frac{1}{m}(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)})g(\theta^T x)(1-g(\theta^T x))\frac{\partial \theta^T x}{\partial \theta_i} \\
&= -\frac{1}{m}(y - g(\theta^T x))x_i \\
\frac{\partial^2 J(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{1}{m}x_i\frac{\partial g(\theta^T x)}{\partial \theta_j} \\
&= \frac{1}{m}x_i x_j g(\theta^T x)(1-g(\theta^T x)) \\
H &= XX^T g(\theta^T x)(1-g(\theta^T x))
\end{aligned}
$$

So, $z^T Hz = (z^T x)^2 g(\theta^T x)(1-g(\theta^T x))$ As we known, $g(\theta^T x)(1-g(\theta^T x))$ is a scalar which > 0, $z^T x$ is also a scalar. then we can conclude $z^T Hz \leq 0$

1.2 b

see *p01b_logreg.py* for detail

1.3 c