CS229 Fall 2017, Problem Set #1: Supervised Learning

Armand Sumo – armandsumo@gmail.com

April 27, 2021

Collaborators:

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

---

1. Logistic regression

Average empirical loss for logistic regression:

$$J(\theta) = -\tfrac{1}{m}\sum_{i=1}^{m} log(h_\theta(y^i x^i))$$

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

(a)

$$\begin{aligned}
\nabla_\theta J(\theta) &= -\frac{1}{m}\sum_{i=1}^{m} \frac{1}{g(\theta^T y^{(i)} x^{(i)})} \nabla_\theta g(\theta^T y^{(i)} x^{(i)}) \\
&= -\frac{1}{m}\sum_{i=1}^{m} \frac{1}{g(\theta^T y^{(i)} x^{(i)})} y^{(i)} x^{(i)} g(\theta^T y^{(i)} x^{(i)})(1 - g(\theta^T y^{(i)} x^{(i)})) \\
&= -\frac{1}{m}\sum_{i=1}^{m} \frac{1}{y}^{(i)} x^{(i)}(1 - g(\theta^T y^{(i)} x^{(i)}))
\end{aligned}$$

$$\begin{aligned}
H_{i,j} = \frac{\partial}{\partial \theta_j}\left[\nabla_\theta J(\theta)\right]_i &= \frac{1}{m}\sum_{i=1}^{m}(y^{(i)})^2 x_j^{(i)} x_i^{(i)} g(\theta^T y^{(i)} x^{(i)})(1 - g(\theta^T y^{(i)} x^{(i)})) \\
&= \frac{\partial}{\partial \theta_j}\left[\nabla_\theta J(\theta)\right]_i \qquad \text{H is symmetric}
\end{aligned}$$

Let's show that for any vector z,
$z^THz \geq 0$

$$\sum_i \sum_j z_i x_i x_j z_j = \sum_i z_i x_i \sum_j z_j x_j = (x^T z)(x^T z) = (x^T z)^2 \geq 0$$

$$\begin{aligned}
z^THz &= \sum_i z_i^T (Hz)_i = \sum_i \sum_j z_i (H_{i,j}) z_j \\
&= \sum_i \sum_j z_i (\frac{1}{m} \sum_{k=1}^{m} (y^{(k)})^2 x_j^{(k)} x_i^{(k)} g(\theta^T y^{(k)} x^{(k)})(1 - g(\theta^T y^{(k)} x^{(k)}))) z_j \\
&= \frac{1}{m} \sum_{k=1}^{m} \sum_i \sum_j (y^{(k)})^2 z_j x_j^{(k)} z_i x_i^{(k)} g(\theta^T y^{(k)} x^{(k)})(1 - g(\theta^T y^{(k)} x^{(k)})) \\
&= \frac{1}{m} \sum_{k=1}^{m} (y^{(k)})^2 g(\theta^T y^{(k)} x^{(k)})(1 - g(\theta^T y^{(k)} x^{(k)}))((x^{(k)})^T z)^2
\end{aligned}$$

For any vector z, $g(z) \in [0, 1]$, hence $z^THz \geq 0$.
This implies that H is positive semi-definite, therefore J is convex and has no local minima other than the global one.

(b) next

Problem 2

(a)