CS229: Problem Set #1

Andrew Ng

0130

Problem 1

1.

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} - \frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)} \log(1 - h_\theta(x^{(i)})))$$

$$= -\frac{1}{m} \sum_{i=1}^{m} [y^{(i)} \frac{1}{h_\theta(x^{(i)})} - (1 - y^{(i)}) \frac{1}{1 - h_\theta(x^{(i)})}] \frac{\partial}{\partial \theta_j} h_\theta(x^{(i)})$$

Then we calculate:

$$\frac{\partial}{\partial \theta_j} h_\theta(x^{(i)}) = \frac{\partial}{\partial \theta_j} g(\theta^T x^{(i)}) = g(\theta^T x^{(i)})(1 - g(\theta^T x^{(i)}))x_j^{(i)} = h_\theta(x^{(i)})(1 - h_\theta(x^{(i)}))x_j^{(i)}$$

And we can further simplify the above equation:

$$\frac{\partial}{\partial \theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [(y^{(i)} - h_\theta(x^{(i)})) - (1 - y^{(i)})h_\theta(x^{(i)})]x_j^{(i)}$$

$$= -\frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$$

Then we can get second-order derivative:

$$H_{ij} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} J(\theta) = \frac{1}{m} \sum_{k=1}^{m} x_j \frac{\partial}{\partial \theta_i} h_\theta(x^{(k)})$$

$$= \frac{1}{m} \sum_{k=1}^{m} x_i^{(k)} x_j^{(k)} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))$$

for each vector z, consider the quadratic form of Hessian matrix:

$$z^T H z = \sum_{i=1}^{n} \sum_{j=1}^{n} z_i H_{ij} z_j = \frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{m} z_i x_i^{(k)} z_j x_j^{(k)} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))$$

$$= \frac{1}{m} \sum_{k=1}^{m} (\sum_{i=1}^{n} \sum_{j=1}^{n} z_i x_i^{(k)} z_j x_j^{(k)}) h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))$$

$$= \frac{1}{m} \sum_{k=1}^{m} (x^{(k)T} z)^2 h_\theta(x^{(k)})(1 - h_\theta(x^{(k)})) \geq 0 \Leftrightarrow H \succeq 0$$

2. Codes are shown in src director, see `src/p01b_logreg.py`.

3.

$$P(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{P(x|y; \phi, \mu_0, \mu_1, \Sigma)}{P(x|y = 1; \phi, \mu_0, \mu_1, \Sigma)P(y = 1) + P(x|y = 0; \phi, \mu_0, \mu_1, \Sigma)P(y = 0)}$$

$$= \frac{\exp(-1/2(x - \mu_1)^T \Sigma^{-1}(x - \mu_1))}{\exp(-1/2(x - \mu_1)^T \Sigma^{-1}(x - \mu_1))\phi + \exp(-1/2(x - \mu_0)^T \Sigma^{-1}(x - \mu_0))(1 - \phi)}$$

$$= \frac{1}{1 + ((1 - \phi)/\phi) \exp(-1/2(x - \mu_0)^T \Sigma^{-1}(x - \mu_0) + 1/2(x - \mu_1)^T \Sigma^{-1}(x - \mu_1))}$$

then we calculate the portion of the denominator inside the exponential term using the fact that: (1) Σ is symmetric (2) if A is symmetric, then $(A^{-1})^T = (A^T)^{-1}$.

$$1/2((x - \mu_1)^T \Sigma^{-1}(x - \mu_1) - (x - \mu_0)^T \Sigma^{-1}(x - \mu_0))$$

$$= (\mu_0 - \mu_1)^T \Sigma^{-1} x + 1/2(\mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1)$$

2

and this allow us to simplify the first equation and prove that the decision boundary is linear:

$$P(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{1}{1 + \exp(-(\theta^T x + \theta_0))}$$

$$\theta = \Sigma^{-1}(\mu_0 - \mu_1)$$

$$\theta_0 = 1/2(\mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1) + \log(1 - \phi) - \log \phi$$

Problem 2

Problem 3

1.

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

$$= \exp(-\lambda + y\log\lambda - \log y!) = \frac{1}{y!}(\log\lambda y - \lambda)$$

$$\Rightarrow b(y) = \frac{1}{y!}, \ \eta = \log\lambda, \ T(y) = y, \ \alpha(\eta) = \lambda = e^\eta$$

2. Let canonical response function represented as g, then we have:

$$\lambda = g(\eta) = e^\eta$$

then we know $g = \exp$.

3.

$$\log p(y|\lambda) = -\lambda + y\log\lambda - \log y! \Rightarrow \log p(y^{(i)}|x^{(i)}; \theta)$$

$$= -e^{\theta^T x} + y^{(i)}\theta^T x^{(i)} - \log y^{(i)}!$$

and we can calculate the derivative of $\log p(y^{(i)}|x^{(i)}; \theta)$ with respect to θ_j:

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

we then get the gradient ascent update rules as follows:

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

In fact, the member in GLM has similar stochastic gradient ascent update rules:

$$\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)} = \theta_j + \alpha(y^{(i)} - \mathbb{E}(y^{(i)}|x^{(i)}; \theta))x_j^{(i)}$$

since $\mathbb{E}(y|x; \theta) = \exp(\theta^T x)$, we get the same answer.

4. Codes are shown in src directory, see `src/p03d_poisson.py`.

Problem 4

1. From the property of the probability space we get:

$$\int_\Omega p(y;\eta)dy = 1 = \frac{1}{\exp(\alpha(\eta))}\int_\Omega b(y)\exp(\eta y)dy$$

which is equivalent to:

$$\exp(\alpha(\eta)) = \int_\Omega b(y)\exp(\eta y)dy$$

apply the partial derivatives to both sides, we have:

$$\frac{\partial}{\partial\eta}\exp(\alpha(\eta)) = \exp(\alpha(\eta))\frac{\partial}{\partial\eta}\alpha(\eta) = \frac{\partial}{\partial\eta}\int_\Omega b(y)\exp(\eta y)dy = \int_\Omega b(y)\exp(\eta y)ydy$$

after some algebraic manipulation, we get:

$$\frac{\partial}{\partial\eta}\alpha(\eta) = \int_\Omega b(y)y\exp(\eta y - \alpha(\eta))dy = \mathbb{E}[Y|X;\theta]$$

2.

$$\frac{\partial^2}{\partial\eta^2}\alpha(\eta) = \frac{\partial}{\partial\eta}(\frac{\partial}{\partial\eta}\alpha(\eta)) = \frac{1}{\exp(\alpha(\eta))}\int_\Omega b(y)y^2\exp(\eta y)dy - \frac{1}{\exp(\alpha(\eta))}\frac{\partial}{\partial\eta}\alpha(\eta)\int_\Omega b(y)y\exp(\eta y)dy$$
$$= \mathbb{E}[Y^2|X;\theta] - \mathbb{E}[Y|X;\theta]^2 = \text{Var}[Y|X;\theta]$$

3. We can formulate the loss function as follows:

$$l(\theta) = -\log J(\theta) = -\log P(Y|X;\theta)$$

where $J(\theta)$ is the likelihood function. In order to get the hessian of the loss function, we first calculate the first-order derivative of the loss function:

$$\frac{\partial}{\partial\theta_j}l(\theta) = \frac{-1}{p(y;\eta)}\frac{\partial}{\partial\eta}p(y;\eta)\frac{\partial}{\partial\theta_j}\eta = \frac{-x_j}{p(y;\eta)}\frac{\partial}{\partial\eta}p(y;\eta)$$

then we calculate the second-order derivative of the loss function:

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}l(\theta) = \frac{x_j}{p(y;\eta)^2}\frac{\partial}{\partial\eta}p(y;\eta)\frac{\partial}{\partial\theta_i}\eta\frac{\partial}{\partial\eta}p(y;\eta) - \frac{x_j}{p(y;\eta)}\frac{\partial^2}{\partial\eta^2}p(y;\eta)\frac{\partial}{\partial\theta_i}\eta$$
$$= \frac{x_ix_j}{p(y;\eta)^2}(\frac{\partial}{\partial\eta}p(y;\eta))^2 - \frac{x_ix_j}{p(y;\eta)}\frac{\partial^2}{\partial\eta^2}p(y;\eta)$$

by calculating the first-order and second-order derivatives of $p(y;\eta)$, we have:

$$\frac{\partial}{\partial\eta}p(y;\eta) = b(y)\exp(\eta y - \alpha(\eta))(y - \frac{\partial}{\partial\eta}\alpha(\eta)) = p(y;\eta)(y - \frac{\partial}{\partial\eta}\alpha(\eta))$$
$$\frac{\partial^2}{\partial\eta^2}p(y;\eta) = p(y;\eta)(y - \frac{\partial}{\partial\eta}\alpha(\eta))^2 - p(y;\eta)\frac{\partial^2}{\partial\eta^2}\alpha(\eta)$$

then we can further simplify the second-order derivative of the loss function by some algebraic manipulation, and the result is:

$$\frac{\partial^2}{\partial\theta_i\theta_j}l(\theta) = x_ix_j\text{Var}[Y|X;\theta]$$

consider the quadratic form of the hessian matrix, we have:

$$z^THz = \text{Var}[Y|X;\theta]\sum_i\sum_j x_ix_jz_iz_j = \text{Var}[Y|X;\theta](z^Tx) \geq 0, \forall z \in \mathbb{R}^n$$

Problem 5