Problem Set #1.

3. (c)

For a training set $\{(x^{(i)}, y^{(i)}); i=1,\dots,m\}$, let the log-likelihood of an example be $\log p(y^{(i)}|x^{(i)};\theta)$.
By taking the derivative of the log-likelihood with respect to θ_j, derive the stochastic gradient ascent update rule for learning using a GLM model with Poisson responses y and canonical response function

Generalized Linear Model

make update right away with each example it looks at.

$i=1 \sim m$

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

<exponential family>

$$p(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta)) \qquad \eta = \theta^T x$$

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

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

$$\log p(y^{(i)}|x^{(i)};\theta) = \log\left\{\frac{1}{y!}\exp((\log\lambda)y - \lambda)\right\}$$

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

$$\lambda = e^{\theta^T x} \qquad = \log\frac{1}{y!} + y \times (\log\lambda) - \lambda$$

$$= \log\frac{1}{y!} + y \times \log \times e^{\theta^T x^{(i)}} - e^{\theta^T x^{(i)}}$$

$$\frac{\partial \log p(y^{(i)}|x^{(i)};\theta)}{\partial\theta_j} = \frac{\partial}{\partial\theta_j}\left\{\log\frac{1}{y!} + y \times \log \times e^{\theta^T x^{(i)}} - e^{\theta^T x^{(i)}}\right\}$$

$$= \frac{\partial}{\partial\theta_j}\left\{y^{(i)} \times \log \times e^{\theta^T x^{(i)}} - e^{\theta^T x^{(i)}}\right\}$$

$$= \frac{\partial}{\partial\theta_j}\left\{y^{(i)} \times \theta^T x^{(i)} - e^{\theta^T x^{(i)}}\right\}$$

$$= y^{(i)} \times x_j^{(i)} - e^{\theta^T x^{(i)}} \cdot x_j^{(i)}$$

$$= x_j^{(i)}\left\{y^{(i)} - e^{\theta^T x^{(i)}}\right\}$$

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