

3. [25 points] Poisson Regression

- (a) [5 points] Consider the Poisson distribution parameterized by λ :

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

Show that the Poisson distribution is in the exponential family, and clearly state the values for $b(y)$, η , $T(y)$, and $a(\eta)$.

- (b) [3 points] Consider performing regression using a GLM model with a Poisson response variable. What is the canonical response function for the family? (You may use the fact that a Poisson random variable with parameter λ has mean λ .)
- (c) [7 points] 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 the canonical response function.
- (d) [7 points] **Coding problem.** Consider a website that wants to predict its daily traffic. The website owners have collected a dataset of past traffic to their website, along with some features which they think are useful in predicting the number of visitors per day. The dataset is split into train/valid/test sets and follows the same format as Datasets 1-3:

`data/ds4_{train,valid}.csv`

We will apply Poisson regression to model the number of visitors per day. Note that applying Poisson regression in particular assumes that the data follows a Poisson distribution whose natural parameter is a linear combination of the input features (*i.e.*, $\eta = \theta^T x$). In `src/p03d_poisson.py`, implement Poisson regression for this dataset and use gradient ascent to maximize the log-likelihood of θ .