CS109A Introduction to Data Science:

Homework 6 AC 209 : GLMs

Harvard University
Fall 2018
Instructors: Pavlos Protopapas, Kevin Rader

Names of people you have worked with goes here:

In [ ]:

```

```

In [1]:

```
# RUN THIS CELL FOR FORMAT
import requests
from IPython.core.display import HTML
styles = requests.get("https://raw.githubusercontent.com/Harvard-IACS/2018-CS109A/ma
HTML(styles)
```

Out[1]:

In [2]:

```
# Imports
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_regression
from scipy.optimize import minimize

%matplotlib inline
```

Question 5 [25 pts]

The probability density function of Poisson distribution is given by:

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

5.1 Show that Poisson distribution belongs to the general exponential distribution family with probability density:

$$f(y|\theta) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right).$$

5.2 For the Poisson distribution calculate the canonical parameter θ in terms of λ, the cumulant function $b(\theta)$, the dispersion parameter ϕ, and the normalization function $c(y, \phi)$.

5.3 Show that the canonical link $g(.)$ for the Poisson regression model is given by:

$$\eta = g(\lambda) = \log(\lambda).$$

5.4 Derive to the *normal equations* that maximizes the likelihood in the Poisson regression model.

Answers

5.1 Show that Poisson distribution belongs to the general exponential distribution family with probability density:

$$f(y|\theta) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right).$$

your answer here We can rewrite the poisson pdf as

$$\begin{aligned} p(y|\lambda) &= \frac{\lambda^y}{y!}e^{-\lambda} \\ &= \frac{e^{yln(\lambda)}}{e^{ln(y!)}}e^{-\lambda} \\ &= \exp(y\ln(\lambda) - \lambda - \ln(y!)) \end{aligned}$$

If we let

$$\begin{aligned} \theta &= \ln(\lambda) \\ \phi &= 1 \\ b(\theta) &= e^{\theta} \\ c(y, \phi) &= -\ln(y!) \end{aligned}$$

Then

$$\exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right) = \exp(y\ln(\lambda) - \lambda - \ln(y!))$$

as required.

5.2 For the Poisson distribution calculate the canonical parameter θ in terms of λ, the cumulant function $b(\theta)$, the dispersion parameter ϕ, and the normalization function $c(y, \phi)$.

your answer here

In terms of λ:

$$\theta = \ln(\lambda)$$

In terms of $b(\theta)$:

$$\theta = \ln(b(\theta))$$

θ is not a function of ϕ or $c(y, \phi)$.

5.3 Show that the canonical link $g(.)$ for the Poisson regression model is given by:

$$\eta = g(\lambda) = \log(\lambda).$$

your answer here

The canonical link is defined such that $g(\lambda) = \theta$. Hence, since $\theta = \ln(\lambda)$, we know that $g(\lambda) = \ln(\lambda)$

5.4 Derive to the *normal equations* that maximizes the likelihood in the Poisson regression model.

your answer here

The normal equations for the exponential family with a canonical link is

$$\sum_{i=1}^{n} \omega_i (y_i - \mu_i) x_i^T = 0$$

where ω_i is the weight of the dispersion parameters, which we assume to be 1, and $g(\mu_i) = x_i^T \beta$ (only for canonical links).

For the Poisson regression with the canonical link, $\mu_i = e^{x_i^T \beta}$ so the normal equaion is

$$\sum_{i=1}^{n} (y_i - e^{x_i^T \beta}) x_i^T = 0$$

In [ ]: