Machine Learning and Statistical Theory II
Generalized Linear Models

Daniel Yu

September 18,2024

Contents

1 Generalized Linear Models

Note: Generalized Linear Models and Splines can be grouped and used together

1.1 Regression

Recall 2 forms of regression:

1. Linear Regression: $(y \in \mathbb{R})$ $Y = \vec{B}^T \vec{X} + \varepsilon$ where $\varepsilon \sim Normal\left(0, \sigma^2\right)$
2. Logistic Regresion:$(y \in \{0, 1\})$ $\log \frac{P(Y=1|X)}{1-P(Y=1|X)} = \vec{B}^T \vec{X}$

Definition 1. A **transformation (link** is defined as follows. There exists a $g(u)$, the link, where $u = E(Y|\vec{X})$ such that while the expected value may not be linear, some function of it is linear, i.e.:

$$g(E[Y|\vec{X}]) = g(\vec{X}) = \vec{B}^T X.$$

We need the link function to be invertible

Remark. This can be thought of as analogous to having **basis** elements which could be anything but still form a vector space where the outputs are linear in regards to these potentially non-linear basis elements. For example, consider the vector space of polynomials of degree n.

Note. Example
linear regression

$$g(u) = I \cdot u = u.$$

logistic regression:

$$g(u) = \log(\frac{u}{1-u}) = \vec{B}^T X.$$

The above is a linear equation in terms of B_i, X_i, even though **probability p and X are nonlinear**

Note. GLM can be viewed as addressing different distributions

1. linear regression - Assume $Y|\vec{X} \sim Normal(\mu = \vec{B}^T X, \sigma^2)$
2. logistic regression - Assume $Y|\vec{X} \sim Ber(p)$
3. Poission ???

Definition 2. Generalized Linear Models
GLM is a flexible extension of ordinary linear regression that allows for response variables (dependent variables) that have error distribution models other than a normal distribution. GLMs consist of three components:

Random Component: $Y|X \sim$ some distribution. (In practice, GLM work particularly well with exponential family of distributions

2. Linear Assumption: Assume there is a linear predictor $\vec{\xi} = \vec{B}^T X$
3. Link: Between random and covariates $\vec{X}$:

$$g(u(\vec{X})) = g(E(Y|\vec{X})) = \vec{\xi} = \vec{B}^T X.$$

and we want g to be invertible

Note. Exponential Family
The exponential family of distributions (works best for GLM):

1. Gaussian
2. Bernoulli

3. Binomial
4. Multinomial
5. Possion
6. Exponential
7. Gamma
8. Laplace
9. Beta
10. etc.

Stuff like **student t, mixture, some uniform distributions** are not exponential

2 Exponential Family

Definition 3. A pdf of a distribution in d-params with the following form is a d-param exponential family density:

$$\begin{aligned} p(\vec{y};\vec{n}) &= \frac{1}{Z(\vec{n})} h(\vec{y}) \exp[\vec{n}T(\vec{y})] \\ &= h(\vec{y}) \exp[\vec{n}T(y) - A(\vec{n})]. \end{aligned}$$

where $A(\vec{n}) = \log Z(\vec{n})$

1. $\vec{n} \in \mathbb{R}^d$ is a natural param of distribution
2. $T(\vec{y}) \in \mathbb{R}^d$ is vector of sufficient statistics (usually $T(\vec{y}) = \vec{y}$). For example, for normal this would be a vector of sample mean and variance
3. $h(\vec{y})$ is underlying measure (usually $h(\vec{y}) = 1$)
4. $A(\vec{n}) = \log Z(\vec{n})$ is the log normalizer, exists to make sure integral of pdf = 1

$$A(\vec{n}) = \log \int h(y) \exp(\vec{n}^T T(\vec{y})) dy.$$

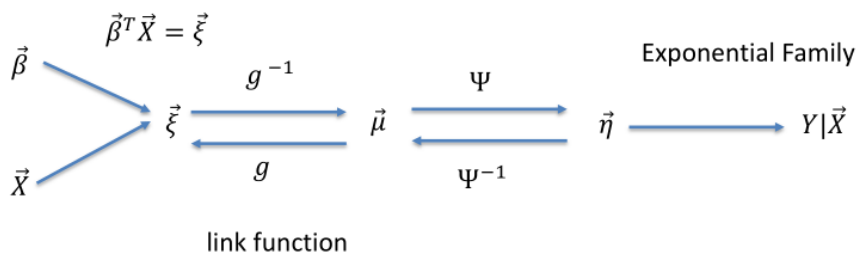

Figure 1: Relationship between Link and Exponential Family Distribution

Note that $\vec{u} = E(Y|X = \vec{x})$. Usually we assume $\xi = \vec{n}$ and $g = \phi$, the **canonical link function**, in this case both sides become symmetric. Note that in this class we will be working with canonical link case. For canonical link:

1. Normal: $g(\mu) = \mu$
2. Binomial: $g(\mu) = \log \frac{\mu}{1-\mu}$ (binary classification)
3. poisson: $g(\mu) = \log(\mu)$
4. gamma: $g(\mu) = -\frac{1}{\mu}$

5. negative binomial: $g(\mu) = \log[\frac{\mu}{k(1+\frac{\mu}{k}}]$

Note. Example - Bernoulli
Normal pdf: $p(y : u) = Ber(y : \mu) = \mu^y(1-\mu)^{1-y}$ where $y \in \{0, 1\}$

$$\begin{aligned} p(y : u) &= \mu^y(1-\mu)^{1-y} \\ &= \exp(y\log(\mu) + (1-y)\log(1-\mu)) \\ &= \exp(y\log(\frac{\mu}{1-\mu}) + \log(1-\mu)). \end{aligned}$$

where $T(y) = y$, $h(\vec{y}) = 1$, $n = \log(\frac{\mu}{1-\mu})$, and $A(\vec{n}) = -\log(1-\mu)$.

Notice that this gives the link function $\mu = \frac{1}{1+e^{-n}}$

TODO Binomial