GENERALIZED LINEAR MODEL

Dinh Huu Nguyen, 03/17/2022

Abstract: an exposition on generalized linear model, how it

- unifies linear regression, logistic regression, Poisson regression, etc.
- frames different loss functions as likelihood
- frames different regularizers as prior

Contents

1. Introduction

A generalize linear model consists of

(1) independent variable X and dependent variable $Y \equiv E(\theta)$ in the exponential family
(2) linear predictor $\nu = \beta X$
(3) link $E(Y \mid X) = g^{-1}(\nu)$ for some choice g (hence the name "link function" for g)

1.1. Likelihood without prior.

For each (x, y)

$$p(Y = y \mid X = x) = p_{Y \mid X=x}(y) \text{ for discreet } Y$$

and

$$p(Y = y \mid X = x) = f_{Y \mid X=x}(y) \text{ for continuous } Y$$

For $(x_1, y_1), \ldots, (x_n, y_n)$

1.2. Likelihood with prior.

2. Examples

2.1. When Y is Gaussian. When $Y \sim N(\mu, \sigma)$ and identity link function $g(t) = t$ then

$$\begin{aligned}
E(Y \mid X = x) &= \mu \\
&= \beta x \\
p(Y = y_j \mid X = x_j) &\sim \frac{e^{-\frac{1}{2}\left(\frac{y_j - \beta x_j}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \\
p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) &= \prod_{j=1}^{n} \frac{e^{-\frac{1}{2}\left(\frac{y_j - \beta x_j}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}
\end{aligned}$$

So to maximize the likelihood on the left hand side, one can minimize its negative log

$$\begin{aligned}
-ln\left(\prod_{j=1}^{n} \frac{e^{-\frac{1}{2}\left(\frac{y_j - \beta x_j}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}\right) &= -\sum_{j=1}^{n} ln\left(\frac{e^{-\frac{1}{2}\left(\frac{y_j - \beta x_j}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}\right) \\
&= \frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - \beta x_j)^2 + c
\end{aligned}$$

that contains the often seen squared error $\sum_{j=1}^{n}(y_j - \beta x_j)^2$.

2.2. When Y is Bernoulli. When $Y \sim Bernoulli(p)$ with logit link function $g(t) = ln\left(\frac{t}{1-t}\right)$ and its inverse $g^{-1}(t) = \frac{e^t}{1+e^t}$ then

$$\begin{aligned}
E(Y \mid X = x) &= p \\
&= \frac{e^{\beta x}}{1 + e^{\beta x}} \\
p(Y = y_j \mid X = x_j) &= p_j^{y_j}(1 - p_j)^{(1-y_j)} \\
&= \left(\frac{e^{\beta x_j}}{1 + e^{\beta x_j}}\right)^{y_j}\left(1 - \frac{e^{\beta x_j}}{1 + e^{\beta x_j}}\right)^{1-y_j} \\
p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) &= \prod_{j=1}^{n}\left(\frac{e^{\beta x_j}}{1 + e^{\beta x_j}}\right)^{y_j}\left(1 - \frac{e^{\beta x_j}}{1 + e^{\beta x_j}}\right)^{1-y_j}
\end{aligned}$$

Again to maximize the likelihood on the left hand side, one can minimize its negative log

$$
\begin{aligned}
-ln\left(\prod_{j=1}^{n}\left(\frac{e^{\beta x_j}}{1+e^{\beta x_j}}\right)^{y_j}\left(1-\frac{e^{\beta x_j}}{1+e^{\beta x_j}}\right)^{1-y_j}\right) &= -\sum_{j=1}^{n} y_j ln\left(\frac{e^{\beta x_j}}{1+e^{\beta x_j}}\right)+(1-y_j)ln\left(1-\frac{e^{\beta x_j}}{1+e^{\beta x_j}}\right)\\
&= -\sum_{j=1}^{n} y_j(\beta x_j)+ln\left(\frac{1}{1+e^{\beta x_j}}\right)\\
&= \sum_{j=1}^{n} ln(1+e^{\beta x_j})-y_j\beta x_j
\end{aligned}
$$

often seen.

2.3. When Y is Poisson.

3. Link with cross entropy

References