Lecture 4 | Generalized linear models

Date de création	@December 23, 2024 5:09 PM
Étiquettes	

▼ Vocab

- **GLMs**: It's a more generic OLS
- natural parameter: **The set of values of η for which the function. is integrable** = canonical parameter
- sufficient statistic: In statistics, **sufficiency** is a property of a statistic computed on a sample dataset in relation to a parametric model of the dataset.
- log partition function: **ore logs based on the routing filter that you configure for each partition in the Log Partitions page**
- Bernoulli: The ***Bernoulli distribution*** is a special case of the binomial distribution where a single trial is conducted (so n would be 1 for such a binomial distribution).
- Response Variable: **Also known as the dependent or outcome variable**, its value is predicted or its variation is explained by the explanatory variable; in an experimental study, this is the outcome that is measured following manipulation of the explanatory variable.

- **Canonical Link Function**: In generalized linear models (GLMs), the canonical link function connects the linear predictor (\\(\eta \\)) to the mean of the response variable (\\(\mu \\)). It is derived from the natural parameter of the exponential family distribution. For example, the canonical link for the normal distribution is the identity function, and for the binomial distribution, it's the logit function.

- **Canonical Response Function**: The canonical response function is the inverse of the canonical link function. It maps the linear predictor η back to the expected value of the response variable μ. For example, for the logit link (used in logistic regression), the response function is the logistic function $\mu = \frac{1}{1+e^{-\eta}}$.
- **Softmax regression**: Softmax regression is a generalization of logistic regression for multi-class classification problems. It models the probabilities of each class as a function of the input features.

- What are the key ideas?
- What terms or ideas are new to me?
- How would I define them?
- How do the ideas relate to what I already know?
- Should not assume that the authors are always correct. Asking appropriate questions.
- What are the good ideas?
- Do the ideias have other applications?

Topics

- Perceptron
- Exponential Family
- GLMs
- 2Softmax Regression (Multiclass Classification)

Notes

Perceptron isn’t that userfull, we apply just for historical reasons.

We can understand Perceptron as a more simple version of logist regression.

$h(x) = g(\Theta^t x)$

$g(z)$ = {1 if $z > 0$, 0 if $z < 0$}

As we saw before, for Logit we have:

$g(z)$= $\frac{1}{1+e^{-z}}$

$h\theta(x)$ = $\frac{1}{1+e^{-(\theta^t x)}}$

Perceptron and Logit have similiar **update rules,** execpt for the fact that $h\theta(x)$ means different things for each.

$$\Theta := \Theta + \alpha(y^i - h\theta(x)^i)xi^j$$

We know that $(y^i - h\theta(x)^i)$ is scalar, it could assume the fallowing values:

- 0 (0 - 0 or 1 - (+1))
- 1 (0 - (-1))
- -1 (0 - (+1))

The Perceptron has a great limitation, it can classify the XOR problem.

Exponentially Family

PDF → Probability Density Function

$p(y; \eta)$ = $b(y)exp(\eta T(y) - \alpha(\eta))$

PDF → Probabillity Of An Event You Want Happens

- To understand GLMs we need understand first Exponential Family.
- Exponential has a PDF in the form $b(y)exp(\eta T(y) - \alpha(\eta))$

- y → data
- η→ natural parameter (control the shape)
- $T(y)$→ Sufficient Statistic (Summarize the data)
- $b(y)$→ Base mesure (help scale the data)
- $a(\eta)$ → log-partition (check for normalization)

This equation model our y data. It help us to make predictions.

Seems like, you can take a distribution like Bernoulli and adjust to seems like our equations for Expontential Familly, right?

Expontential Family ≠ from Exponential Distribution, some are:

- Gaussian
- Poisson
- Gamma, Exponential
- Bernoulli
- Beta, Dirichlet (Basian Statistics)

Depending on you problem, we use differente exponential member, for exemple:

- regression → gaussian
- binary classification → bernoulli

GLMs

A GLMs are extensions of the exponential family to include our X (features).

In order to distiguised GMLs from Exponential Family we need make some assumptions:

1. $y|x;\theta$ ~ exponential family (η)
2. $\eta = \Theta^T x$
3. Test Time: output $E[y|x;\theta]$
    a. this means that our $h\theta(x)$ = $E[y|x;\theta]$

For our members in the Exponential Family The **Learning Updatating** *Rule* is the same. Which is $\Theta_i := \Theta_j + \alpha(y^i - h_\theta(x^i))x^i_j$

Terminolagy

$\mu = E[y;\eta] = g(\eta)$ → Cononical Response Function

$\eta = g^{-1}(\mu)$ → "…" Link Function

and $g(\eta) = \frac{\partial a(\eta)}{\partial \eta}$

Logistic Regression

$h_\theta(x) = E[y|x_i\theta] = \phi$ = $\frac{1}{1+e^{-\eta}}$ = $\frac{1}{1+e^{-\theta^T x}}$

> The type of distribution depends of your problem (regression requires guassian, while binary classficiation bernoulli).

> GMLs is just a generla way to modeling data.

Soft Max Classification

We can use Softmax classification for Multiclassficiation problems, you can approach softmax as a GLM but this isn't mandatory.

$cross(P, P_hat) = -\sum_y \in (o, a, b, c)\ p(y) P_hat(y)$

we treat this like loss and apply gradient descent in respect to Θ.

Questions

- What the update rule seems equal for Logit, Linear Regression and Perceptron?
- What's the purpose of $b(y)exp(\eta T(y) - \alpha(\eta))$?

  This equation model our y data. It help us to make predictions.
- Why is this userful, transform a distribution (e.g Bernoulli, Poisson) to Exponterial Family?

  It simplifies analysis, modeling and the computacional resources use.
- What's PDF?

  a mathematical function that describes the probability of different possible values of a variable.

- Why we use expontential families?

  It has some nice mathemetical propreties

  - What propeties?
    a. MLE with respect to η in concave
       a. NLL is concave
    b. $E(y; \eta) = \frac{\partial a(\eta)}{\partial \eta}$, this mean: this is the mean of distribution
    c. $Var(y; \eta) = \frac{\partial^2 a(\eta)}{\partial \eta^2}$

    b and **c** are nice becuase you generally need to integrate to get mean and variance, but, in this case is just defferentiation.

▼ What's the types of parametrization?

- model: θ
- natural: η
- canonical: ϕ → bernoulli; $\mu\delta^2$ → gaussian; λ → Poisson

$\Theta^T x$ give you η (model → natural) and between the two are g(natural → canonical) and g^{-1} (natural → model)