Lecture 4 Perceptron & Generalized Linear Model

1. Exponential Family

$$p(y;\eta) = \underbrace{b(y)}_{\text{base measure}} \exp(\boxed{\eta}^T \underline{T(y)} - \boxed{a(\eta)})$$

sufficient statistics

Log partition function

natural parameter (canonical parameter)

* $\exp(-a(\eta))$: normalization constant - intergrates y to 1

by vary η, we can get different distributions

Real - Gaussian;
Binary - Bernoulli,
Count - Poisson,
$\mathbb{R}^+$ - Gamma, Exponential,
Distribution - Beta, ...

Properties:

① MLE with respect to η is concave.
NLL (Negative Log Likelihood) is convex

② $E[y;\eta] = \frac{\partial}{\partial\eta} a(\eta)$

③ $Var[y;\eta] = \frac{\partial^2}{\partial\eta^2} a(\eta)$

2. GLM

Assumptions:

① $y|x;\theta \sim$ Exponential Family

② $\eta = \theta^T x$

③ Given x, goal is to predict $T(y)$ (mostly $T(y)=y$)

$\Rightarrow h(x) = E[y|x]$

3. GLM Training

The learning update rule: $\theta_j := \theta_j + \alpha\left(\sum_{i=1}^{m}(y^{(i)} - h_\theta(x^{(i)}))\, x_j^{(i)}\right)$

Terminology: η - natural parameter

$\mu = E[y;\eta] = g(\eta) \longrightarrow$ canonical response function $\quad g(\eta) = \frac{\partial}{\partial\eta}a(\eta)$

$\eta = g^{-1}(\mu) \rightarrow$ canonical link function

3 parameters:

Model Parameter　　　natural Parameter　　　canonical parameter

$$\theta \qquad\qquad \eta \qquad\qquad \phi\text{ - Bernoulli};\ \mu\,\sigma^2\text{ - Gaussian};\ \lambda\text{ - Poisson}$$

$$\underset{\text{learn}}{\uparrow} \qquad \xrightarrow[\text{Design choice}]{\theta^T x} \qquad \underset{\xleftarrow{g^{-1}}}{\xrightarrow{g}}$$

e.g. Logistic Regression

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

4. Softmax Regression - Multi-class classification (Multinomial Logistic Regression)

$$h_\theta(x) = \begin{bmatrix} P(y=1|x;\theta) \\ \vdots \\ P(y=k|x;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} \exp(\theta^{(j)T} x)} \begin{bmatrix} \exp(\theta^{(1)T} x) \\ \vdots \\ \exp(\theta^{(k)T} x) \end{bmatrix}$$

cost function: $J(\theta) = -\left[\sum_{i=1}^{m} (1-y^{(i)}) \log(1-h_\theta(x^{(i)})) + y^{(i)} \log h_\theta(x^{(i)})\right]$

$$= -\left[\sum_{i=1}^{m}\sum_{k=0}^{1} 1\{y^{(i)}=k\}\log P(y^{(i)}=k \mid x^{(i)};\theta)\right]$$

$$= -\left[\sum_{i=1}^{m}\sum_{k=1}^{K} 1\{y^{(i)}=k\}\log \frac{\exp(\theta^{(k)T}x^{(i)})}{\sum_{j=1}^{K}\exp(\theta^{(j)T}x^{(i)})}\right]$$