CMPUT 367

Roderick Lan

Contents

1 Midterm Exam **2**

2 Exponential Family **2**

2.1 Linear Regression 2

2.2 Logistic Regression 3

3 Nonlinear Models **3**

Lecture 10 - Feb 13

1 Midterm Exam

3 questions; up to Feb 15 lecture

---

2 Exponential Family

(EXP)
(rev from last lecture)

$$p(x;\eta) = h(x)\exp\{\eta^\top T(x) - A(\eta)\}$$
$$A(\eta) = \ln \int h(x)\exp(\eta^\top T(x))dx$$

$A(\eta)$ is convex
log likelihood is convex in η
moment generating property

$$\frac{\partial A(\eta)}{\partial \eta} = \mathop{\mathbb{E}}_{x\sim p(x;\eta)}[T(x)] \qquad \frac{\partial^2 A(\eta)}{\partial \eta\, \partial \eta^\top} = \mathop{\mathrm{Cov}}_{x\sim p(x;\eta)}[T(x)]$$

MLE $\iff$ Moment Matching
organize moment params - same as moment stat (?; 27 min feb 13 lec)
$\hat{\mu} = \mathop{\mathbb{E}}_{x\sim p(x;\eta)}[T(x)]$

2.1 Linear Regression

$$\{(x^{(m)}, t^{(m)})\}_{m=1}^M$$

$t^{(m)}|x^{(m)} \sim \mathrm{EXP}(\eta^{(m)})$
with canonical response function

$$\eta^{(m)} = w^\top x^{(m)}$$

GLM:

$$t^{(m)}|x^{(m)} \sim \mathrm{EXP}(w^\top x^{(m)})$$

$w^\top x^{(m)}$ are the natural parameters

$$\ln p(t^{(1)}\cdots t^{(m)}|x^{(1)}\cdots x^{(m)}) = \sum_{m=1}^M \ln \cdot h(x^{(m)})\exp(w^\top x^{(m)} t - A(\eta))$$

$$\begin{aligned}\frac{\partial}{\partial w}\ln p &= \sum_{m=1}^M \left[t^{(m)}\cdot x_i^{(m)} - \frac{\partial A(\eta)}{\partial \eta}\cdot\frac{\partial \eta}{\partial w_i}\right] \\ &= \sum_{m=1}^M \left[t^{(m)}\cdot x_i^{(m)} - \mathop{\mathbb{E}[t^{(m)}]}_{t^{(m)}\sim \mathrm{EXP}(t^{(m)}|x^{(m)};\eta)}\cdot x_i^{(m)}\right]\end{aligned}$$

2.2 Logistic Regression

$$\eta^{(m)} = w^\top x^{(m)}$$

mean param from natural param; need ψ inverse

$$\mu^{(m)} = \psi^{-1}(\eta^{(m)}) = \sigma(\eta^{(m)})$$

Linear poisson regression (see feb 13 notes)
(Canonical response func chosen bc gradient is nice; prediction uses mean params instead of natural)

3 Nonlinear Models

Design nonlinear features
Nonlinear kernels (68 min)

> if model only depends on
> $$[x^{(i)}]^\top [x^{(j)}]$$
> we may extened the notion of inner product $\langle x^{(i)}, x^{(j)} \rangle$

Neural networks - a stack of linear predictions (learnable)

Non linear problems exist