CMPUT 367

Roderick Lan

Contents

1 Optima **2**
1.1 Propositions 3

2 Exp Family, GLM **4**

Lecture 8 - Feb 6

1 Optima

Definition 1.0.1: Global Optimum (min)

x is a global optimum of f iff $\forall y \in \text{dom}_f,\ f(y) \geq f(x)$

Definition 1.0.2: Local Optimum (min)

x is a local optimum of f iff
$\exists \epsilon > 0$, $\forall y \in \text{dom}_f$ if $\|y - x\| < \epsilon$ then $f(y) \geq f(x)$
or
$\forall y$ in the neighborhood of x $(N_\epsilon(x))$, $f(y) \geq f(x)$

Thm 1.0.1

Let f be a convex function
A local optimum of f, x, is a global optimum

There must exist some ϵ s.t. $\forall z \in N_\epsilon(x), f(z) \geq f(x)$
Fix such ϵ for further use

Consider any $y \in \text{dom}_f$

Case 1: $\|y - x\| < \epsilon$
$f(y) \geq f(x)$ already known by local optimality

Case 2: $\|y - x\| \geq \epsilon$
(z should be halfway b/w x and ϵ boundary)
Pick $\lambda = \dfrac{\|y - x\| - \frac{1}{2}\epsilon}{\|y - x\|}$

$$z = \lambda x + (1 - \lambda) y$$

Then,[1]

$$\begin{aligned}
\|z - x\| &= \left\| \frac{\|y - x\| - \frac{1}{2}\epsilon}{\|y - x\|} x + \frac{\frac{1}{2}\epsilon}{\|y - x\|} y - x \right\| \\
&= \left\| \frac{-\frac{1}{2}\epsilon}{\|y - x\|} x + \frac{\frac{1}{2}\epsilon}{\|y - x\|} y \right\| = \frac{1}{2}\epsilon
\end{aligned}$$

$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) \geq f(x)$
$\lambda f(x) + (1 - \lambda) f(y) \geq f(x)$
$f(y) \geq f(x)$

[1] 28 min for feb 6 Lecture

1.1 Propositions

1. If f is a convex function, $\nabla f(x)$ exists at x_0, and $\nabla f(x_0) = 0$
   Then x is a local/global optimum

2. For a differentiable function f, the gradient descent

   $$w^{(\text{new})} = w^{(\text{old})} - \alpha \nabla f(w^{(\text{old})})$$

   If $\nabla f(w^{(\text{old})}) \neq 0$ and $\alpha > 0$ is small enough, then

   $$f(w^{(\text{new})}) < f(w^{(\text{old})})$$

Proof 1.1

Prove via taylor expansion.

$$\begin{aligned} f(w^{(\text{new})}) &= f(w^{(\text{old})}) + [\nabla f(w^{(\text{old})})]^\top (w^{(\text{new})} - w^{(\text{old})}) + \text{h.o.t} \\ &= f(w^{(\text{old})}) + [\nabla f(w^{(\text{old})})]^\top (-\alpha \nabla f(w^{\text{old}})) \end{aligned}$$

∇f is a descending direction
$M \nabla f(x)$ if $M \geq 0$[a]

[a] 52 min in feb 6 lecture

2 Exp Family, GLM

Discussed topics: Linear, Logistic, Softmax
All of them have probabilistic assumption $\Rightarrow$ Loss function $\Rightarrow$ Optimization

$$(y_i - t_i)x_i$$

Exponential Family and Generalized Linear models

Definition 2.0.1: Exponential Family

$x \sim \text{EXP}(\eta)$ if

$$p(x;\eta) = h(x)\exp(\eta^\top T(x) - A(\eta))$$

$T(x)$ - sufficient statistics; features of x
η - natural parameter
$h(x)$ - unparameterized residual [a]
$A(\eta)$ - normalization term

$h(x)$ can be controlled (ie. how residual calculated)
$T(x)$ can be controlled[b]

[a]62 min feb 6 lec
[b]66 min feb 6 lec

$$\int h(x)\exp(\eta^\top T(x) - A(\eta))dx = 1$$

$$\int h(x)\frac{\exp(\eta^\top T(x))}{\exp(A(\eta))}dx = 1$$

$$A(\eta) = \ln \int h(x)\exp(\eta^\top T(x))dx$$

"Logged Linear Model"

Example 2.0.1

$$\begin{aligned}
p(x_i;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}x^2 + \frac{1}{\sigma^2}x\mu - \frac{1}{2\sigma^2}\mu^2\right\} \\
&= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}x^2 + \frac{1}{\sigma^2}x\mu - \frac{1}{2\sigma^2}\mu^2 + \ln\frac{1}{\sigma}\right\} \\
&= h(x)\exp\{\eta^\top T(x) - A(\eta)\}
\end{aligned}$$

σ cant be in $h(x)$ since $h(x)$ must be unparameterized

$$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

The quantities that involve the params μ and σ^2

$$\eta = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix}$$

Example 2.0.2: Bernoulli

$x \sim \text{Bernoulli}(\pi)$
$x \sim \pi\langle 1\rangle + (1-\pi)\langle 0\rangle$

$$p(x;\pi) = \pi^{\mathbb{1}\{x=1\}} \cdot (1-\pi)^{\mathbb{1}\{x=0\}}$$

Gaussian (linear) and Bernoulli (classification) distribution both part of exponential function family