CMPUT 367

Roderick Lan

Contents

Lecture 7 - Feb 1

1 Convexity Cont'd

Thm 1.0.1

Let f be a twice differentiable function on a convex domain;
the following three conditions are equivalent

1. f is a convex function
2. [1st Order] $\forall x, y \in \text{dom} f$,

$$f(y) \geq f(x) + [\nabla f(x)]^\top (y - x)$$

   (Linear approx is always a lower bound)

3. 2nd order $\forall x \in \text{dom} f$

$$\nabla^2 f(x) \geq 0$$

   (positive semidefinite)

1D: $f''(x) \geq 0$ then curve is convex
High-D:

> if $\nabla^2 f(x)$ diagonal - eigenval are diagonal elements;
> nonnegative evals $\rightarrow$ every dimension is curving up $\rightarrow$ convex
>
> $\nabla^2 f(x)$ not diagonal - check eigenvals;
> negative eval $\rightarrow$ not convex
> nonnegative eval only $\rightarrow$ convex

Proof 1.1: 2. $\rightarrow$ 1.

$\forall x, y \in \mathrm{dom} f;\quad \forall \lambda \in (0,1)$
The first order condition:

$$\forall x, y \in \mathrm{dom} f : f(y) \geq f(x) + (\nabla_x f(x))^\top (y - x)$$

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

$$f(y) \geq f(z) + [\nabla_z f(z)]^\top (y - z) \tag{1}$$

Put z into x and x into y

$$f(x) \geq f(z) + [\nabla_z f(z)]^\top (x - z) \tag{2}$$

$\lambda(2) + (1-\lambda)(1)$ gives us:

$$\begin{aligned}
\lambda f(x) + (1-\lambda) f(y) &\geq \lambda f(z) + (1-\lambda) f(z) \\
&\quad + (1-\lambda)[\nabla_z f(z)]^\top (y - z) + \lambda [\nabla_z f(z)]^\top (x - z) \\
&\geq f(z) + [\nabla_z f(z)]^\top (y - z - \lambda y + \lambda z + \lambda x - \lambda z) \\
&= f(z) + [\nabla_z f(z)]^\top [\lambda x + (1-\lambda) y - z]
\end{aligned}$$

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

$$= f(z) + [\nabla_z f(z)]^\top [0] = f(z)$$

Proof 1.2: 3. $\rightarrow$ 2.

$\forall x, y \in \mathrm{dom} f$

$$f(y) = f(x) + [\nabla f(x)]^\top (y - x) + \frac{1}{2}(y - x)^\top [\nabla^2 f(z)](y - x)$$

for some z between x and y.
Taylor Expansion to first order:

$$T(y) = f(x) + [\nabla f(z)]^\top (y - x) \text{ for } z$$

(MVT) Since $\nabla^2 f(z) \geq 0$, second order term ≥ 0; Thus

$$f(y) \geq T(y)$$

2 MSE

$$\nabla_w J_{\text{MSE}} = \frac{1}{M}(X^\top X w - Xt)$$
$$\nabla w_w^2 J_{\text{MSE}} = \frac{1}{M} X^\top X$$

$\forall v, v^\top X^\top X v = (Xv)^\top X v = \|Xv\|_2^2 \geq 0$

Definition 2.0.1: Optimality

x is a **Global Optimum** for f iff

$$\forall y \in \text{dom} f,\ f(x) \leq f(y)$$

x is a **Local Optimum** of f iff

$$\exists \epsilon > 0 \text{ s.t. } \forall y \in \text{dom} f$$

if $\|x - y\| < \epsilon$ then $f(x) \leq f(y)$

Thm 2.0.1

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