STAT 538 Lecture 6

Convex functions

©Marina Meilă
mmp@stat.washington.edu

Reading: BV 3.1–3.3. These notes are supplements to the reading.

1 Additional examples of convex functions

(All functions f, g below are assumed to be convex.)

Generic functions

- $h(x, y) = f(x)g(y)$ is convex.
  Proof
  $$\nabla^2 h = \begin{bmatrix} \nabla^2 f & 0 \\ 0 & \nabla^2 g \end{bmatrix}$$

Functions from statistics

- The normalization constant of an exponential family $Z(\theta)$
  Proof $e^{-\theta x}$ is convex as a function of $\theta \in \mathbb{R}$ for any x; $e^{-\sum_i \theta_i x_i}$ is also convex in $[\theta_1, \ldots \theta_n]$ as a product of convex functions of different variables. Then $Z(\theta) = \sum_x e^{-\theta^T x}$ is convex as a sum of convex functions.
- $\log Z(\theta)$ is convex
  Proof The proof is statistical. Essentially, one can show that
  $$E_\theta[x] = \sum_x x p(x) = -\nabla \ln Z(\theta) \qquad Var_\theta(x) = \sum_x x x^T p(x) = \nabla^2 \ln Z(\theta)$$
  The convexity follows from the positive-definiteness of the variance. For first equality:
  $$\sum_x x_i \frac{e^{-\sum_i \theta_i x_i}}{Z} = \sum_x \frac{\frac{\partial e^{-\sum_i \theta_i x_i}}{\partial \theta_i}}{Z} = -\frac{\frac{\partial Z}{\partial \theta_i}}{Z} = -\frac{\partial \ln Z}{\partial \theta_i} \qquad (1)$$

For the second, we write the element i, j of the co-variance matrix

$$
\begin{aligned}
Var(X)_{ij} &= \sum_x x_i x_j \frac{e^{-\sum_l \theta_l x_l}}{Z} - \left(\sum_x x_i \frac{e^{-\sum_l \theta_l x_l}}{Z}\right)\left(\sum_{x'} x'_j \frac{e^{-\sum_l \theta_l x'_l}}{Z}\right) && (2)\\
&= \sum_x \frac{-x_i}{Z^2}\left[-x_j e^{-x^T\theta} Z - (e^{-x^T\theta})\left(\sum_{x'} -x'_j e^{-x^T\theta}\right)\right] && (3)\\
&= \sum_x \frac{\partial}{\partial x_j}\left(\frac{-x_i e^{-x^T\theta}}{Z}\right) = \frac{\partial}{\partial x_j}\frac{\partial \ln Z}{\partial x_j} && (4)
\end{aligned}
$$

- Any marginal of a discrete distribution.
  Let $X, Y \in \Omega_X \times \Omega_Y$ with $\Omega_X| = m$, $|\Omega_Y| = n$ be two discrete random variables, and let Θ be the set of all probability distributions over $\Omega_X \times \Omega_Y$. That is, we define $P_\theta(X = i, Y = j) = \theta_{ij}$; then $\Theta = \{\theta = [\theta_{ij}]_{ij} \in [0,1]^{m\times n}, \sum_{ij} \theta_{ij} = 1\}$ The marginal $P_X(i) = \sum_j \theta_{ij}$ is a linear function of the entries of θ, therefore it is convex.

2 Stricly convex and strongly convex functions

A function is **strictly convex** if Jensen's inequality is strict whenever $t \in (0, 1)$, i.e.

$$
tf(x) + (1-t)f(x') > f(tx + (1-t)x') \quad \text{for all } t \in (0,1) \tag{5}
$$

The concept of subgradient is a generalization of the gradient for functions which are not differentiables. A **subgradient** of a convex function f at point x is any vector $g \in \mathbb{R}^n$ so that

$$
f(x') \geq f(x) + g^T(x' - x) \quad \text{for all } x' \in \text{dom} f \tag{6}
$$

In other words, g is a subgradient iff it is the normal of a supporting hyperplane of the epigraph f at x. It follows immediately that aa convex function admits a subgradient at any point in its domain. [**Exercise:** Show that $\partial f(x) = \{g \in \mathbb{R}^n \,|\, g \,\text{subgradient of}\, f \,\text{at}\, x\}$ is a convex set.] If $\nabla f(x)$ exists, then it is the unique subgradient at x.

A function f is μ-**strongly convex** iff

$$f(x') \geq f(x) + g^T(x' - x) + \frac{\mu}{2}||x' - x||^2 \quad \text{for all } x, x' \in \text{dom} f \text{ and all } g \in \partial f(x) \tag{7}$$

The notion of strong convexity is a generalization of the condition

$$\nabla^2 f(x) \succ \mu I \tag{8}$$

from doubly differentiable functions to all convex functions. [Exercise: Show that (8) implies (7) when the Hessian is defined everywhere.] Strong convexity implies strict convexity, but the converse is not true. For example, the function $f(x) = 1/x$, $x \in (0, \infty)$ is strictly convex but not strongly convex.