Jensen's Inequality

Konstantinos G. Derpanis
kosta@cs.yorku.ca
Version 1.0

March 12, 2005

In this note the concept of *convexity* and *Jensen's Inequality* are reviewed. Jensen's Inequality plays a central role in the derivation of the Expectation Maximization algorithm [1] and the proof of *consistency* of maximum likelihood estimators.

Definition *Let $f(x)$ be a real valued function defined on the interval $I = [a, b]$. f is said to be* **convex** *if for every $x_1, x_2 \in [a, b]$ and $0 \geq \lambda \geq 1$,*

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$$

A function is said to be **strictly convex** *if the equality is strict for $x_1 \neq x_2$.*

Definition *$f(x)$ is said to be concave (strictly concave) if $-f(x)$ is convex (strictly convex).*

Intuitively, the definition of convexity states that function falls below never above the straight line between the points $(x_1, f(x_1))$ to $(x_2, f(x_2))$ (see Fig. 1).

Theorem 0.1 *If $f''(x)$ exists on $[a, b]$ and $f''(x) \geq 0$ on $[a, b]$ then $f(x)$ is convex on $[a, b]$.*

Theorem 0.2 (Jensen's Inequality) *Let $f(x)$ be a convex function defined on an interval I. If $x_1, x_2, \ldots, x_N \in I$ and $\lambda_1, \lambda_2, \ldots, \lambda_N \geq 0$ with $\sum_{i=1}^{N} \lambda_i$,*

$$f\left(\sum_{i=1}^{N} \lambda_i x_i\right) \leq \sum_{i=1}^{N} \lambda_i f(x_i)$$

Alternatively, if $f(x)$ is a convex function and $X \in \{x_i : 1, \ldots, N\}$ is a random variable with probabiliites $P(x_i)$ where $\sum P(x_i) = 1$, then,

$$f(E\{X\}) \leq E\{f(X)\}$$
$$f\left(\sum_{i=1}^{N} x_i P(x_i)\right) \leq \sum_{i=1}^{N} f(x_i) P(x_i)$$

Proof The proof is by induction. For the base case $N = 1$, the theorem is trivially true. When $N = 2$,

$$f(x_1)P(x_1) + x_2 P(x_2)) \leq f(x_1)P(x_1) + f(x_2)P(x_2)$$

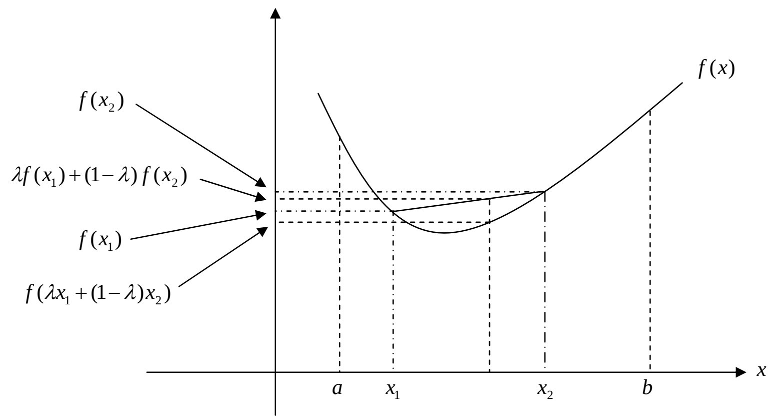

Figure 1: Illustrative example of convexity.

This is true by the definition of convex functions.
Inductive hypothesis: Suppose that the theorem is true for $N = k-1$. Let $P'(x_i) = P(x_i)/(1-P(x_k))$ for $i = 1, 2, \ldots, k-1$,

$$\begin{aligned}
\sum_{i=1}^{k} f(x_i)P(x_i) &= (1 - P(x_k)) \sum_{i=1}^{k-1} f(x_i)P'(x_i) + f(x_k)P(x_k) \\
&\geq (1 - P(x_k)) f\left(\sum_{i=1}^{k-1} x_i P'(x_i)\right) + f(x_k)P(x_k) \text{ (By inductive hypothesis)} \\
&\geq f\left((1 - P(x_k)) \sum_{i=1}^{k-1} x_i P'(x_i) + x_k P(x_k)\right) \text{ (By base case } N = 2\text{)} \\
&= f\left(\sum_{i=1}^{k-1} x_i P(x_i) + x_k P(x_k)\right) \\
&= f\left(\sum_{i=1}^{k} x_i P(x_i)\right)
\end{aligned}$$

Hence, the theorem is true by induction. ∎

Example Since $\ln(x)$ is concave, by Jensen's inequality the following holds,

$$\ln\left(\sum_{i=1}^{N} x_i P(x_i)\right) \geq \sum_{i=1}^{N} \ln(x_i) P(x_i)$$

This result is used in the derivation of the EM algorithm [1].

References

[1] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximal likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39:185–197, 1977.