Information Theory, Inference and Learning Algorithms

David J.C. MacKay

▶ 2.7 Jensen's inequality for convex functions

The words 'convex$\smile$' and 'concave$\frown$' may be pronounced 'convex-smile' and 'concave-frown'. This terminology has useful redundancy: while one may forget which way up 'convex' and 'concave' are, it is harder to confuse a smile with a frown.

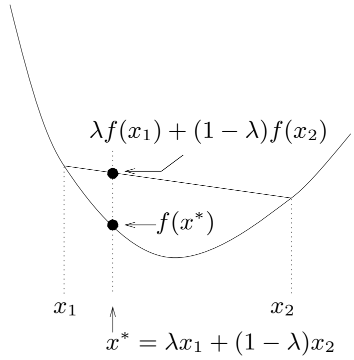


Figure 2.10. Definition of convexity.

Convex$\smile$ functions. A function $f(x)$ is *convex*$\smile$ over (a,b) if every chord of the function lies above the function, as shown in figure 2.10; that is, for all $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \;\; \leq \;\; \lambda f(x_1) + (1-\lambda) f(x_2). \tag{2.47}$$

A function f is *strictly convex*$\smile$ if, for all $x_1, x_2 \in (a,b)$, the equality holds only for $\lambda = 0$ and $\lambda = 1$.

Similar definitions apply to concave$\frown$ and strictly concave$\frown$ functions.

Some strictly convex$\smile$ functions are

- x^2, e^x and e^{-x} for all x;
- $\log(1/x)$ and $x \log x$ for $x > 0$.

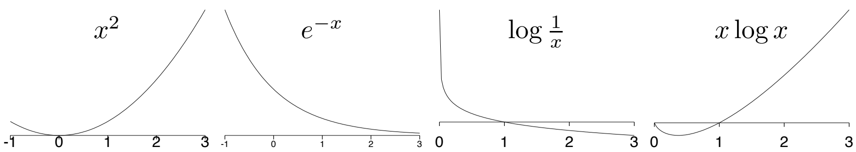


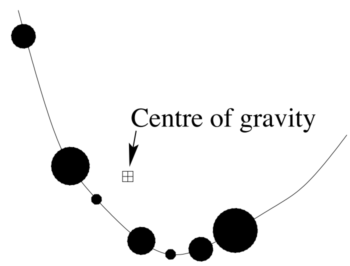
Figure 2.11. Convex$\smile$ functions.

Jensen's inequality. If f is a convex$\smile$ function and x is a random variable then:

$$\mathcal{E}\left[f(x)\right] \geq f(\mathcal{E}[x]), \tag{2.48}$$

where $\mathcal{E}$ denotes expectation. If f is strictly convex$\smile$ and $\mathcal{E}\left[f(x)\right] = f(\mathcal{E}[x])$, then the random variable x is a constant.

Jensen's inequality can also be rewritten for a concave$\frown$ function, with the direction of the inequality reversed.



A physical version of Jensen's inequality runs as follows.

If a collection of masses p_i are placed on a convex$\smile$ curve $f(x)$ at locations $(x_i, f(x_i))$, then the centre of gravity of those masses, which is at $(\mathcal{E}[x], \mathcal{E}\left[f(x)\right])$, lies above the curve.

If this fails to convince you, then feel free to do the following exercise.

Exercise 2.14.[*2*, p.41] Prove Jensen's inequality.

Example 2.15. Three squares have average area $\bar{A} = 100\,\mathrm{m}^2$. The average of the lengths of their sides is $\bar{l} = 10\,\mathrm{m}$. What can be said about the size of the largest of the three squares? [Use Jensen's inequality.]