DIGRESSION:

JENSENS INEQUALITY & CONVEXITY

Definition: A function $f(x)$ is said to be *convex* over an interval (a, b) if for every $x_1, x_2 \in (a, b)$ and $0 \leq \lambda \leq 1$,

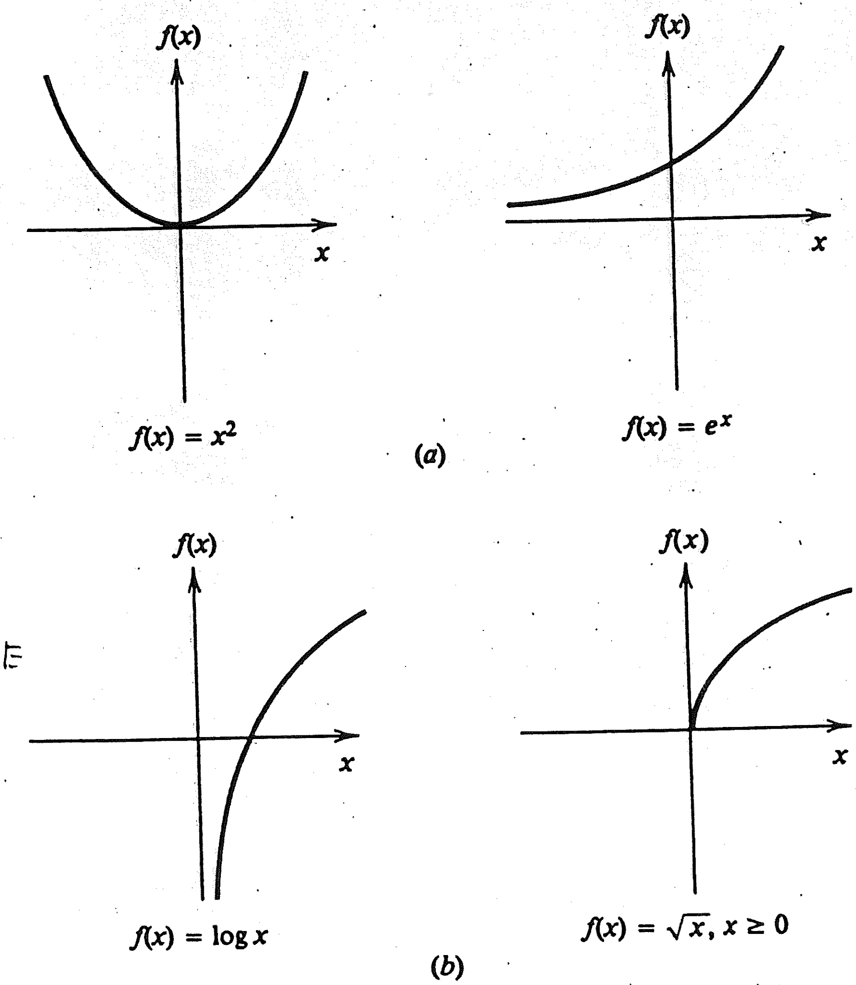
$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2). \tag{2.72}$$

A function f is said to be *strictly convex* if equality holds only if $\lambda = 0$ or $\lambda = 1$.

Definition: A function f is *concave* if $-f$ is convex.

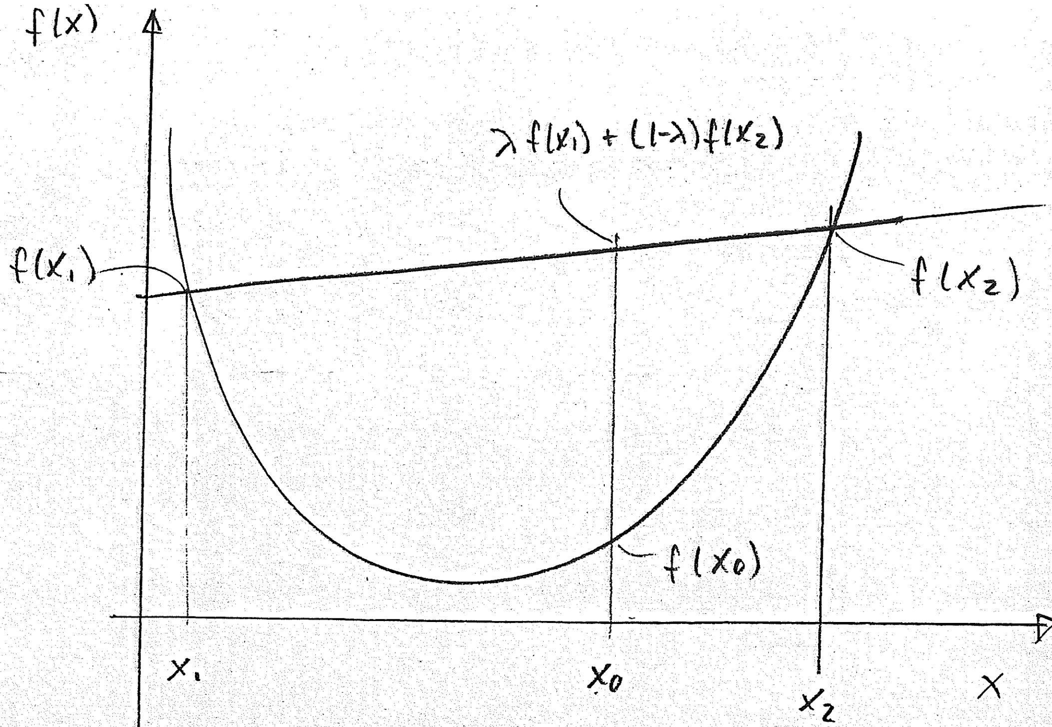
CONVEX

CONCAVE



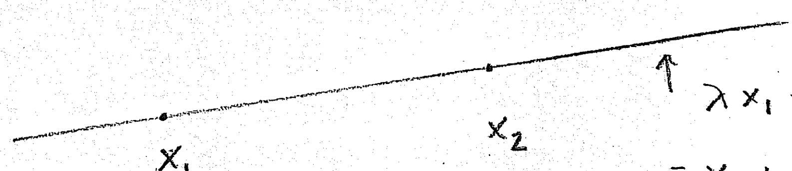
POS. 2ND DERIVATIVE IS SUFFICIENT

NEG. 2ND DERIVATIVE IS SUFFICIENT



$$x_0 = \lambda x_1 + (1-\lambda) x_2$$

$$\forall\ 0 \le \lambda \le 1 : \quad f(\lambda x_1 - (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$



$\uparrow\ \lambda x_1 + (1-\lambda) x_2$

$= x_1 + (1-\lambda)(x_2 - x_1)$

FOR WHOLE LINE $\lambda \in \mathbb{R}$

FOR SEGMENT $\lambda \in [0,1]$

Theorem 2.6.1: *If the function f has a second derivative which is non-negative (positive) everywhere, then the function is convex (strictly convex).*

Proof: We use the Taylor series expansion of the function around x_0, i.e.,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x^*)}{2}(x - x_0)^2 \tag{2.73}$$

where x^* lies between x_0 and x. By hypothesis, $f''(x^*) \geq 0$, and thus the last term is always non-negative for all x.

$$\geq f(x_0) + f'(x_0)(x - x_0)$$

We let $x_0 = \lambda x_1 + (1 - \lambda)x_2$ and take $x = x_1$ to obtain

$$f(x_1) \geq f(x_0) + f'(x_0)[(1 - \lambda)(x_1 - x_2)] . \tag{2.74}$$

Similarly, taking $x = x_2$, we obtain

$$f(x_2) \geq f(x_0) + f'(x_0)[\lambda(x_2 - x_1)] . \tag{2.75}$$

Multiplying (2.74) by λ and (2.75) by $1 - \lambda$ and adding, we obtain (2.72). The proof for strict convexity proceeds along the same lines. □

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

$$\geq \underbrace{(\lambda + 1 - \lambda)}_{1} f(x_0) + \underbrace{\lambda(1-\lambda) f'(x_0)(x_1 - x_2) + \lambda(1-\lambda) f'(x_0)(x_2 - x_1)}_{0}$$

Let E denote expectation. Thus $EX = \sum_{x \in \mathcal{X}} p(x)x$ in the discrete case and $EX = \int xf(x)\,dx$ in the continuous case.

The next inequality is one of the most widely used in mathematics and one that underlies many of the basic results in information theory.

Theorem 2.6.2 (*Jensen's inequality*): *If f is a convex function and X is a random variable, then*

$$Ef(X) \ge f(EX) . \tag{2.76}$$

Moreover, if f is strictly convex, then equality in (2.76) implies that $X = EX$ with probability 1, i.e., X is a constant.

Proof: We prove this for discrete distributions by induction on the number of mass points. The proof of conditions for equality when f is strictly convex will be left to the reader.

For a two mass point distribution, the inequality becomes

$$p_1 f(x_1) + p_2 f(x_2) \ge f(p_1 x_1 + p_2 x_2) , \tag{2.77}$$

which follows directly from the definition of convex functions. Suppose the theorem is true for distributions with $k-1$ mass points. Then writing $p_i' = p_i/(1-p_k)$ for $i = 1, 2, \ldots, k-1$, we have

$$\sum_{i=1}^{k} p_i f(x_i) = p_k f(x_k) + (1-p_k) \sum_{i=1}^{k-1} p_i' f(x_i) \tag{2.78}$$

INDUCTION

$$\ge p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} p_i' x_i \right) \tag{2.79}$$

$$\ge f\left(p_k x_k + (1-p_k) \sum_{i=1}^{k-1} p'_i x_i\right)$$

BINARY CASE

$$= f\left(\sum_{i=1}^{k} p_i x_i\right)$$

CONTINUOUS CASE PROVEN USING CONTINUITY ARGUMENTS !

□