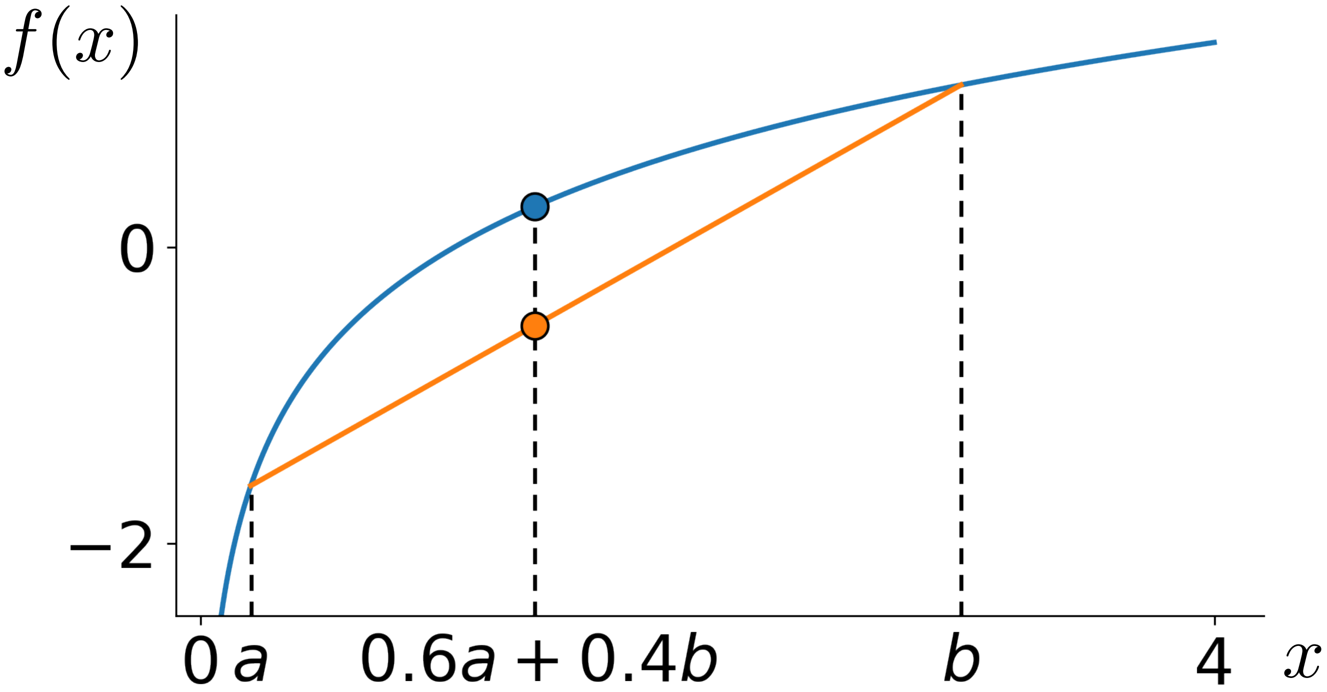
General form of Expectation Maximization

Concave functions



Def.: $f(x)$ is concave if

for any a, b, α : $f(\alpha a + (1-\alpha)b) \geq \alpha f(a) + (1-\alpha) f(b)$

$$0 \leq \alpha \leq 1$$

Jensen's inequality

If we have a CONCAVE function:

If $$f(\alpha a + (1-\alpha)b) \geq \alpha f(a) + (1-\alpha) f(b)$$

Then $$\alpha_1 + \alpha_2 + \alpha_3 = 1;\ \alpha_k \geq 0.$$

NOTE: alpha1, 2, 3 here are the probabilities.

$$f(\underbrace{\alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3}_{\mathbb{E}_{p(t)} t}) \geq \underbrace{\alpha_1 f(a_1) + \alpha_2 f(a_2) + \alpha_3 f(a_3)}_{\mathbb{E}_{p(t)} f(t)}$$

NOTE the expectation here!

$$p(t = a_1) = \alpha_1,$$
$$p(t = a_2) = \alpha_2,$$
$$p(t = a_3) = \alpha_3$$

Jensen's inequality

If $$f(\alpha a + (1-\alpha)b) \geq \alpha f(a) + (1-\alpha)f(b)$$

Then Jensen's inequality:

$$f\left(\mathbb{E}_{p(t)} t\right) \geq \mathbb{E}_{p(t)} f(t)$$

This is jensen's inequality summarized.

HOLDS TRUE for an arbitrary amount of a1, a2, a3, ..., an.

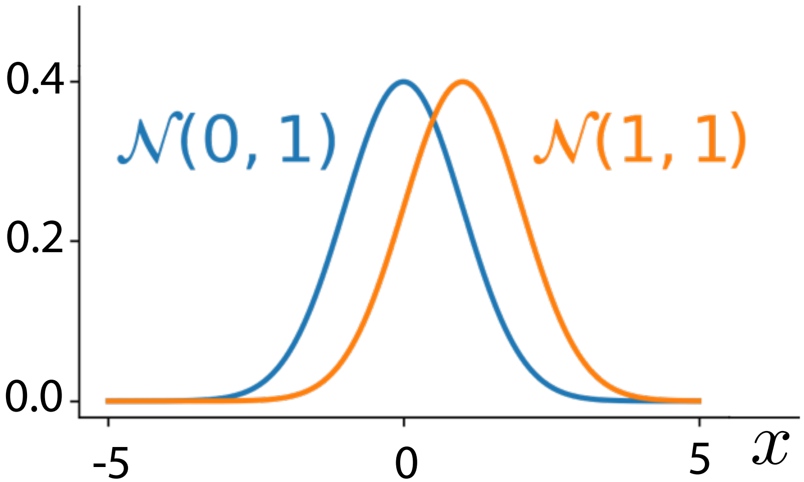
I.e. if the function f is concave , then f of expected value of t is geq expected value of f(t).

Kullback–Leibler divergence

Essentially, we can't measure parameter difference to figure out the 'difference' of two distributions.
E.g. both below are 1 diff but obviously the right difference is lesser than the left difference.

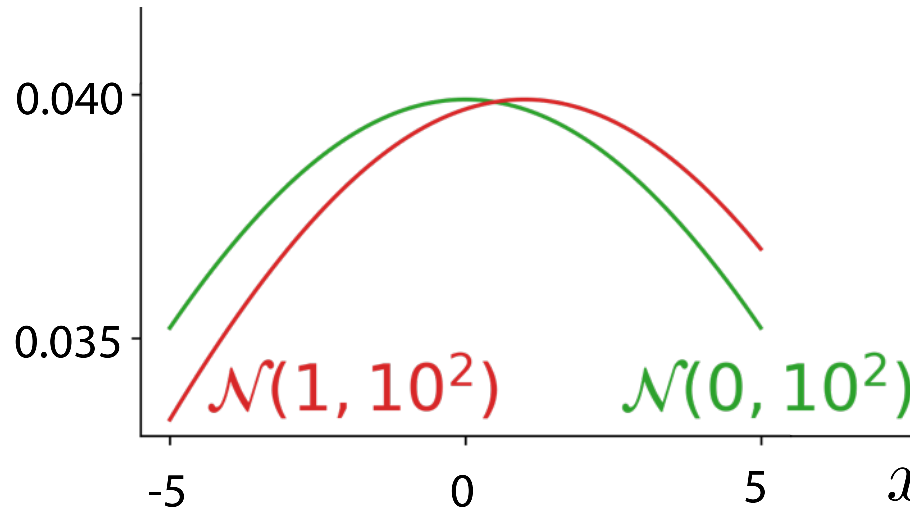
Parameters difference: 1

$$\mathcal{KL}(q_1 \parallel p_1) = 0.5$$



Parameters difference: 1

$$\mathcal{KL}(q_2 \parallel p_2) = 0.005$$



$$\mathcal{KL}(q \parallel p) = \int q(x) \log \frac{q(x)}{p(x)} dx$$

KL(q||p) = E_{x follow q pdf} (log q(x) / p(x))
Boils down to :
KL(q||p) = H(q,p) - H(q)
where H(q,p) is crossentropy, H(q) is simply entropy.

Kullback–Leibler divergence

$$\mathcal{KL}(q \parallel p) = \int q(x) \log \frac{q(x)}{p(x)} dx$$

1. $\mathcal{KL}(q \parallel p) \neq \mathcal{KL}(p \parallel q)$
2. $\mathcal{KL}(q \parallel q) = 0$
3. $\mathcal{KL}(q \parallel p) \geq 0$

Proof: $$-\mathcal{KL}(q \parallel p) = \mathbb{E}_q \left(-\log \frac{q}{p} \right) = \mathbb{E}_q \left(\log \frac{p}{q} \right)$$

$$\leq \log(\mathbb{E}_q \frac{p}{q}) = \log \int q(x) \frac{p(x)}{q(x)} dx = 0$$

Kullback–Leibler divergence

$$\mathcal{KL}(q \parallel p) = \int q(x) \log \frac{q(x)}{p(x)} dx$$

Summary

A way to compare distributions
not a proper distance

1. $\mathcal{KL}(q \parallel p) \neq \mathcal{KL}(p \parallel q)$
2. $\mathcal{KL}(q \parallel q) = 0$
3. $\mathcal{KL}(q \parallel p) \geq 0$