Machine Learning Notes - 2

Lipschitz Continuity

https://math.stackexchange.com/questions/2374289/understanding-lipschitz-continuity/2374305#2374305

Think about the mean value theorem and Lipschitz continuity.
Mean value theorem says if f is continuous at [a,b] and differentiable at (a,b), then

$$\exists c \in (a,b) \text{ such that } \frac{f(b)-f(a)}{b-a} = f'(c).$$

Lipschitz says that

$$\exists K > 0, \forall a, b \in D_f, \text{ such that } \frac{|f(b)-f(a)|}{|b-a|} \leq K.$$

Then if the derivative of f as a function is bounded, then f will be **Lipschitz**.

Specifically, if **K == 1**, then the function is said to be **1-Lipschitz**.

Consider the case

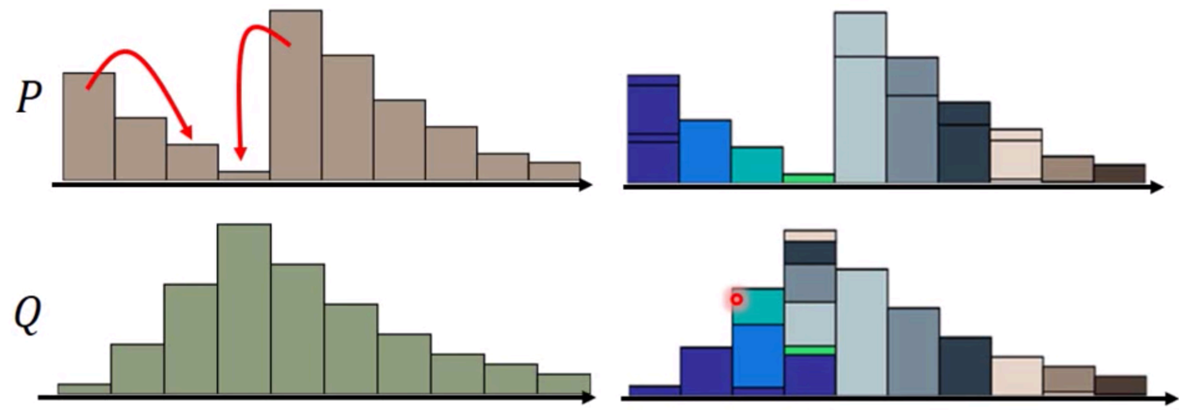
$$f(x) = \sqrt{x} \text{ for } x \in [0, 1]$$

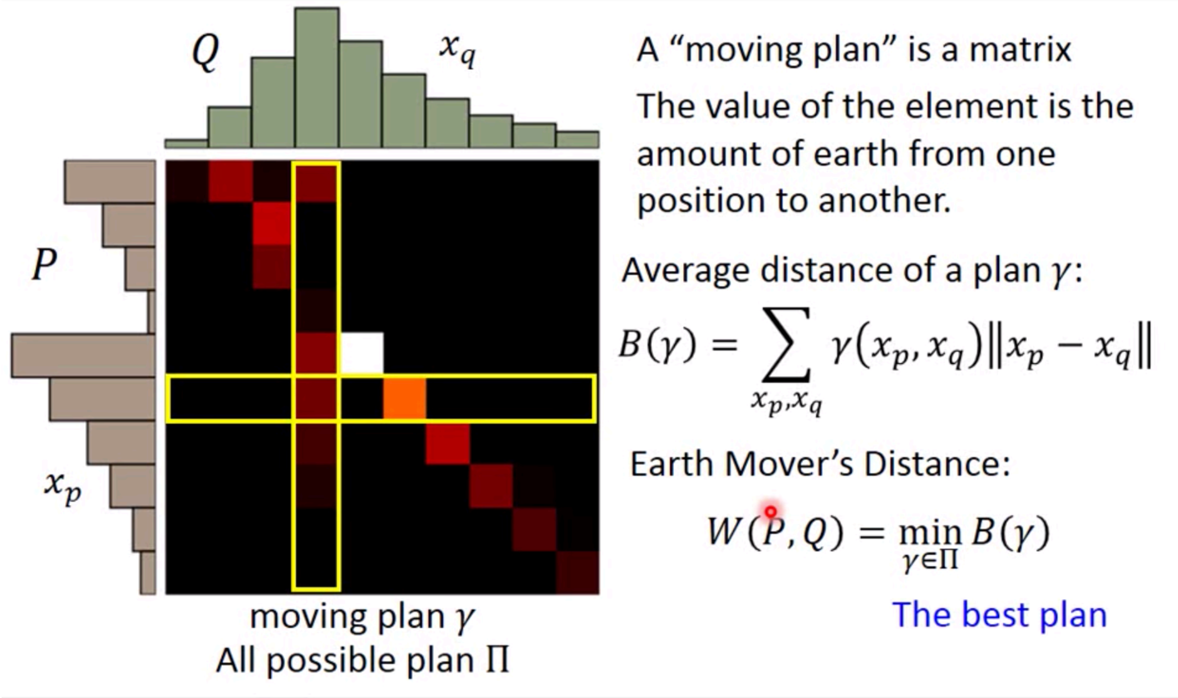
the f is not Lipschitz, since

$$\sup_{x \in [0,1]} f'(x) = \lim_{x \to 0} f'(x) = +\infty.$$

Earth Mover's Distance

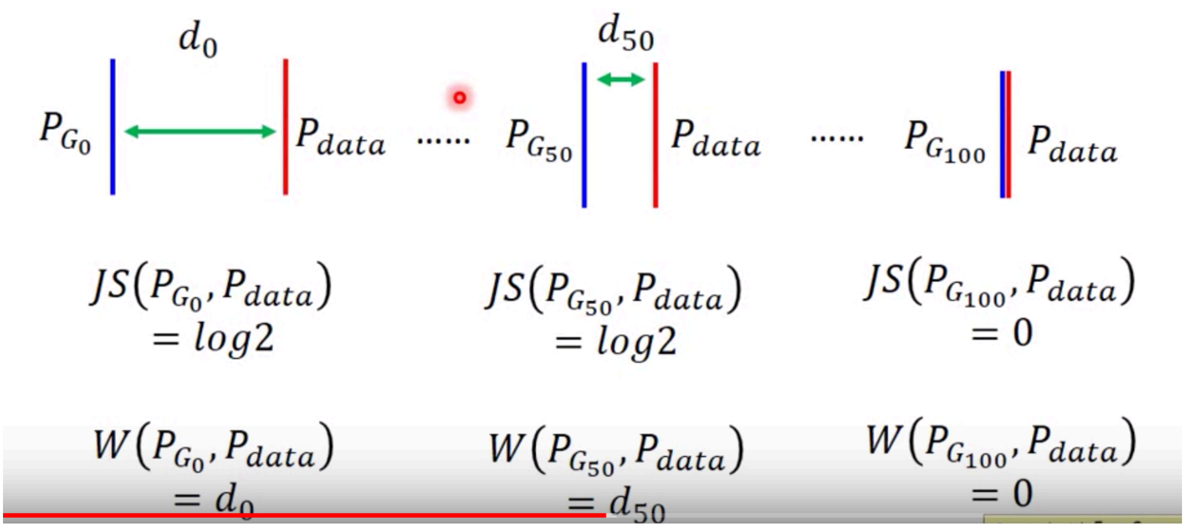
https://www.youtube.com/watch?v=KSN4QYgAtao&feature=youtu.be
Consider one distribution **P as a pile of earth**, and another distribution **Q as the target**,
Earth Mover's Distance is the average distance the earth mover has to move the earth.





Why Earth Mover’s Distance ?

the comparation between **JS-Divergence** and **Earth Mover’s Distance**
P_G denotes the distribution of Generator and **P_{DATA}** denotes the distribution of real data.
As shown below, **JSD** is less effectiveness to represent the “distance” between **P_G** and **P_{DATA}** while **Earth Mover’s Distance** does.



Back to the GAN framework

$$D_f(P_{data}||P_G) \Rightarrow W(P_{data}, P_G)$$

$$= \max_D \{E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}[f^*(D(x))]\}$$

$$W(P_{data}, P_G) = \max_{D\in 1-Lipschitz} \{E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}[D(x)]\}$$