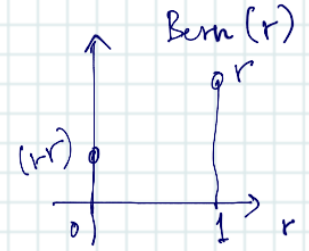
25/8/23

Deep Learning

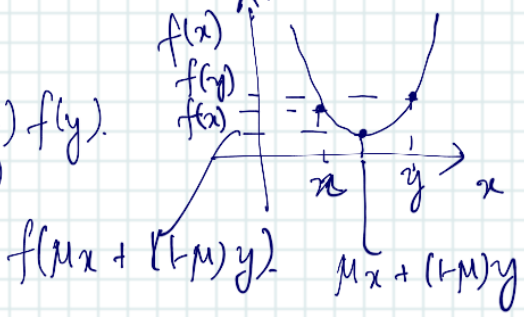
- Recap
- Examples
- Convex function, Jensen's inequality
- $D(p||q) \geq 0$; $D(p||q) \neq D(q||p)$.; Mutual Information
- Jensen Shannon Divergence $JS(p||q)$.
- Cross entropy $H(p,q)$
  - Intuition
- Glimpse of DL applications



- Recap: $H(X) = -\sum_x p(x) \log_2 p_x(x) = \sum_x p(x) \log_2 \frac{1}{p_x(x)}$
  - $H(Y|X) = \sum p(x) \cdot H(Y|x) = E_{xy}\left[\log \frac{1}{p_{y|x}(y|x)}\right]$ (Conditional) $H(X) \geq 0$
  - $H(X,Y) = H(X) + H(Y|X)$. (Joint)
  - $D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$ (Relative entropy or Kullback-Leibler divergence).
  - Show that $D(p||q) \geq 0$, with $D(p||q) = 0$ iff $p = q$.

Convex function: $f: X \to \mathbb{R}$ where X is a convex set and satisfies for all $x, y \in X$, $0 \leq \mu \leq 1$ the property that

$$f(\mu x + (1-\mu) y) \leq \mu f(x) + (1-\mu) f(y) \quad -(1)$$



Jensen's Inequality: If f is a convex function

$$E[f(x)] \geq f[E[x]] \quad -(2)$$

Also note that log is a concave function $\Rightarrow$ the inequalities $\geq \to \leq$ & vice versa in (1) & (2).

Use these hints to show that $D(p||q) \geq 0$.

$$D(p||q) = E_x \log\left[p(x)/q(x)\right] \geq 0.$$

Example: If $X \sim \text{Bern}(r)$ find $H(X)$

$$H(X) = r \cdot \log_2 \frac{1}{r} + (1-r) \cdot \log_2 \frac{1}{(1-r)}$$

Example: a) Find $D(p||q)$ $p \sim \text{Bern}(r)$; $q \sim \text{Bern}(s)$.

b) Find $D(q||p)$

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = r \cdot \log \frac{r}{s} + (1-r) \cdot \log \frac{(1-r)}{(1-s)} \quad - (3)$$

$$D(q||p) = s \cdot \log \frac{s}{r} + (1-s) \cdot \log \frac{(1-s)}{(1-r)} \quad (4)$$

Clearly (3) ≠ (4) ; (3) = (4) when $r = s$.

$D(p||q) = 0$ when $r = s$.

Note: KL divergence is not a valid distance since $D(p||q) \neq D(q||p)$

- Jensen Shannon Divergence:

$$JS(p||q) = \frac{1}{2}\left[D(p||m) + D(q||m) \right]$$

$$m = \frac{1}{2}(p+q)$$

  - Symmetric & bounded.

- Mutual Information: $I(X;Y) = H(X) - H(X|Y)$

- Cross Entropy: $H(p,q) = H(p) + D(p||q)$

  - $H(p,q)$: ground truth PMF, Estimated PMF
  - $H(p)$: Entropy of ground truth RV
  - $D(p||q)$: Extra bits needed to describe p when we observe q