24/8/23

Deep Learning

- Recap
- Probability basics: RV, PMF, Joint PMF, Conditional PMF, Marginal
- Basics of information theory: Entropy $H(X)$, Joint Entropy $H(X,Y)$, Conditional Entropy $H(X|Y)$
- Example: p_{XY}

$Y \backslash X$	1	2	3	4
1	1/8	1/16	1/32	1/32
2	1/16	1/8	1/32	1/32
3	1/16	1/16	1/16	1/16
4	1/4	0	0	0

Recap: Probabilistic setting: Assume that data points are drawn from a fixed but unknown dist.

- RV: $X : \Omega \to \mathbb{R}$ (a measurable function)
- PMF: $p_X : \mathbb{R} \to [0,1]$ (restrict our attention to discrete RVs).
- Joint PMF: $p_{XY} : \mathbb{R} \times \mathbb{R} \to [0,1]$
- Conditional PMF: $p_{Y|X}(y \mid X = x)$
  - Recall the optimal Bayes' classifier: $\boxed{y^* = \underset{y \in \mathcal{X}}{\operatorname{argmax}}\ p(y \mid X = x)}$
- Marginal density: $p_{XY}(x,y)$ and we want to find the marginal $p_X(x)$

$$p_X(x) = \sum_y p_{XY}(x,y)$$

- Example 1: Find $p_X(x)$ and $p_Y(y)$ from $p_{XY}(x,y)$ specified above.
  - $p_X(x) = \sum_{y=1}^{4} p_{XY}(x,y)$

    $p_X(X=1) = \frac{1}{2}$.

    $p_X(X=2) = \frac{1}{4}$

    $p_X(X=3) = \frac{1}{8}$

    $p_X(x=4) = \frac{1}{8}$

  |||ly $p_Y(y) = \sum_x p_{XY}(x,y)$

  $p_Y(Y=1) = p_Y(y=2) = p_Y(Y=3) = p_Y(Y=4) = \frac{1}{4}$.

- Example 2: Find $p_Y(Y|X=1) = \frac{p_{XY}(X=1, Y)}{p(X=1)}$

- $p_Y(Y=1|X=1) = \frac{1}{4}$

$p_Y(Y=2|X=1) = \frac{1}{8}$

$p_Y(Y=3|X=1) = \frac{1}{8}$

$p_Y(Y=4|X=1) = \frac{1}{2}$

- Entropy: The number of bits required to describe a random variable. If X is a discrete RV,

$$H(X) = -\sum p_X(x) \cdot \log_2 p_X(x)$$

Example 3: Find $H(Y)$. $H(Y) = -\sum_{y=1}^{4} p_Y(y) \log_2 p_Y(y)$

$$= -4 \cdot \frac{1}{4} \cdot \log_2 \frac{1}{4}$$

$$= 2 \text{ bits.}$$

Example 4: Find $H(X)$: $H(X) = -\sum_{x=1}^{4} p_X(x) \cdot \log_2 p_X(x)$

$$= -\left[\frac{1}{2} \cdot \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + 2 \cdot \frac{1}{8} \cdot \log \frac{1}{8}\right]$$

$$= -\left[\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \times 3\right]$$

$$= 1.75 \text{ bits}$$

- Show $H(X) \geq 0$

- $H(X) = \sum p_X(x) \log_2 \frac{1}{p_X(x)}$ we know $0 \leq p(x) \leq 1$
$\Rightarrow 1 \leq \frac{1}{p(x)} \leq \infty$ } both non-neg.

- Joint Entropy: $H(X,Y) = \sum_x \sum_y p_{XY}(x,y) \cdot \log_2 \left[\frac{1}{p_{XY}(x,y)}\right]$

$$= E_{XY}\left[\log \frac{1}{p_{XY}(x,y)}\right]$$

Example 5: Find $H(X,Y)$ for the joint PMF specified above.

$$H(X,Y) = \sum_{x=1}^{4} \sum_{y=1}^{4} p_{XY}(x,y) \log_2 \frac{1}{p_{XY}(x,y)}$$

- Conditional Entropy: $H(Y|X) = \sum_x p_X(x) \cdot H(Y|X=x)$

$$= \sum_x p_X(x) \cdot \sum_y p_{Y|X}(y|x) \cdot \log_2 \frac{1}{p_{Y|X}(y|x)}$$

$$= \sum \sum p_{XY}(x,y) \cdot \log_2 \frac{1}{p_{Y|X}(y|x)}$$

$$= E_{XY} \log_2 \frac{1}{p_{Y|X}(y|x)}$$

Example 6: Find $H(Y|X)$ for the above joint PMF.

- $H(X,Y) = H(X) + H(Y|X)$

$$H(X,Y) = \sum_x \sum_y p_{XY}(x,y) \cdot \log_2 \frac{1}{p_{XY}(x,y)}$$

$$= -\sum \sum p_{XY}(x,y) \log_2 \left[p(y|x) \cdot p(x)\right]$$

$$= -\sum\sum p_{XY}(x,y) \cdot \left[\log_2 p(x) + \log_2 p(y|x)\right]$$

$$= H(X) + H(Y|X)$$

- Relative Entropy or Kullback-Leibler divergence.

$$D(p\|q) = \sum_x p(x) \cdot \log \frac{p(x)}{q(x)}$$

"Distance": $D(p\|q) \geq 0$; $D(p\|q) \Leftrightarrow p \equiv q$.

- Example 7: If $X \sim \text{Bern}(r)$, find $H(X)$.
- Example 8: If p is $\text{Bern}(r)$ and q is $\text{Bern}(s)$ find $D(p\|q)$.