Lecture 3: Distributions

Distribution

1. X is a distribution over the sample space $\mathcal{S}$

Distribution

1. X is a distribution over the sample space $\mathcal{S}$
   1. Assigns probability p_s to the element $s \in \mathcal{S}$

Distribution

1. X is a distribution over the sample space $\mathcal{S}$
    1. Assigns probability p_s to the element $s \in \mathcal{S}$
2. Y is a distribution over the same sample space $\mathcal{S}$ and assigns q_s probability to element $s \in \mathcal{S}$

Distribution

1. X is a distribution over the sample space $\mathcal{S}$
    1. Assigns probability p_s to the element $s \in \mathcal{S}$
2. Y is a distribution over the same sample space $\mathcal{S}$ and assigns q_s probability to element $s \in \mathcal{S}$
3. Difference between distributions

Distribution

1. X is a distribution over the sample space $\mathcal{S}$
    1. Assigns probability p_s to the element $s \in \mathcal{S}$
2. Y is a distribution over the same sample space $\mathcal{S}$ and assigns q_s probability to element $s \in \mathcal{S}$
3. Difference between distributions
4. Prediction Advantage

Distribution

1. X is a distribution over the sample space $\mathcal{S}$
   1. Assigns probability p_s to the element $s \in \mathcal{S}$
2. Y is a distribution over the same sample space $\mathcal{S}$ and assigns q_s probability to element $s \in \mathcal{S}$
3. Difference between distributions
4. Prediction Advantage
5. Intuition: More the difference, easier to predict whether the sample was sampled according to the distribution X or Y

Total Variation Distance

Definition (Total Variation Distance)

$$\mathrm{SD}\left(X, Y\right) := \sum_{s \in \mathcal{S}} \left| p_s - q_s \right|$$

Total Variation Distance

Definition (Total Variation Distance)

$$\mathrm{SD}\left(X,Y\right) := \sum_{s\in\mathcal{S}} \left|p_s - q_s\right|$$

1. Intuition?

Total Variation Distance

Definition (Total Variation Distance)

$$\mathrm{SD}(X, Y) := \sum_{s \in \mathcal{S}} |p_s - q_s|$$

1. Intuition?
2. Another definition:

$$\max_{A} |\Pr[s \sim X : A(s) = 1] - \Pr[s \sim Y : A(s) = 1]|$$

Total Variation Distance

Definition (Total Variation Distance)

$$\mathrm{SD}\left(X, Y\right) := \sum_{s \in \mathcal{S}} \left|p_s - q_s\right|$$

1. Intuition?
2. Another definition:

$$\max_{A} \left|\Pr\left[s \sim X : A(s) = 1\right] - \Pr\left[s \sim Y : A(s) = 1\right]\right|$$

3. Equivalent!

Prediction Advantage

- Definition:

$$\max_{A} \Pr\left[b \xleftarrow{\$} \{0,1\}, s \sim (1-b)X + bY : A(s) = b\right] - \frac{1}{2}$$

Prediction Advantage

- Definition:

$$\max_{A} \Pr\left[b \xleftarrow{\$} \{0,1\}, s \sim (1-b)X + bY : A(s) = b\right] - \frac{1}{2}$$

- Relation with Total Variation Distance?

Prediction Advantage

- Definition:

$$\max_{A} \Pr\left[b \xleftarrow{\$} \{0,1\}, s \sim (1-b)X + bY : A(s) = b\right] - \frac{1}{2}$$

- Relation with Total Variation Distance?
- Think: SD between U_n and $f(U_{n-1})$ for any function $f\colon \{0,1\}^{n-1} \to \{0,1\}^n$

Prediction Advantage

- Definition:

$$\max_A \Pr\left[b \xleftarrow{\$} \{0,1\}, s \sim (1-b)X + bY : A(s) = b\right] - \frac{1}{2}$$

- Relation with Total Variation Distance?
- Think: SD between U_n and $f(U_{n-1})$ for any function $f\colon \{0,1\}^{n-1} \to \{0,1\}^n$
- Think: What if there are more than two distributions?

Ensembles of Probability Distributions

Definition

A sequence $\{X_n\}_{n\in\mathbb{N}}$ is called an ensemble if for each $n \in \mathbb{N}$, X_n is a probability distribution over $\{0,1\}^*$.

Ensembles of Probability Distributions

Definition

A sequence $\{X_n\}_{n\in\mathbb{N}}$ is called an ensemble if for each $n \in \mathbb{N}$, X_n is a probability distribution over $\{0,1\}^*$.

- Generally, X_n will be a distribution over the sample space $\{0,1\}^{\ell(n)}$ (where $\ell(\cdot)$ is a polynomial)

Ensembles of Probability Distributions

Definition

A sequence $\{X_n\}_{n\in\mathbb{N}}$ is called an ensemble if for each $n \in \mathbb{N}$, X_n is a probability distribution over $\{0,1\}^*$.

- Generally, X_n will be a distribution over the sample space $\{0,1\}^{\ell(n)}$ (where $\ell(\cdot)$ is a polynomial)
- How to measure difference between two ensembles?

Ensembles of Probability Distributions

Definition

A sequence $\{X_n\}_{n \in \mathbb{N}}$ is called an ensemble if for each $n \in \mathbb{N}$, X_n is a probability distribution over $\{0,1\}^*$.

- Generally, X_n will be a distribution over the sample space $\{0,1\}^{\ell(n)}$ (where $\ell(\cdot)$ is a polynomial)
- How to measure difference between two ensembles?
  - Distance

Ensembles of Probability Distributions

Definition

A sequence $\{X_n\}_{n \in \mathbb{N}}$ is called an ensemble if for each $n \in \mathbb{N}$, X_n is a probability distribution over $\{0,1\}^*$.

- Generally, X_n will be a distribution over the sample space $\{0,1\}^{\ell(n)}$ (where $\ell(\cdot)$ is a polynomial)
- How to measure difference between two ensembles?
  - Distance
  - Prediction Advantage