Probability Cheet Sheet

Peitian Zhang

June 30, 2021

Contents

1 Distribution

A distribution $P : \Omega \to \Gamma$ is a mapping from the probabilities of events to the probabilities of outcomes. **A distribution is defined by a *probability mass function* (discrete random variable) or a *probability density function* (continuous random variable)**.

1.1 Joint Distribution

Say a distribution is defined over multiple **discrete** random variables, formally

$$p(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

then $p(X_1, X_2, \cdots, X_n)$ is called **joint distribution**.

2 Bayesian Theorem

Bayesian Theorem is used to infer the *unobserved variable* from the *observed data*, formally

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \tag{1}$$

where $y \in \mathcal{Y}$ is the observed data and $x \in \mathcal{X}$ is the latent variable, $\mathcal{Y}$ denotes the data(event) space and $\mathcal{X}$ denotes the latent space,

- $p(x|y)$ describes the probability of x given y observed, namely the **posterior**.
- $p(y|x)$ describes the probability of y given a determined x, which is the principle that we assume the data is generated, namely the **likelihood**.
- $p(x)$ describes the probability of x regardless of any observed data y, namely the **prior**.
- $p(y)$ describes the probability of y independent of any latent variable, namely the **evidence**.

Furthermore,

$$p(y) = \begin{cases} \int_{x \in \mathcal{X}} p(y|x)p(x)\,dx & X \text{ is continuous,} \\ \sum_{x \in \mathcal{X}} p(y|x)p(x) & X \text{ is discrete.} \end{cases} \tag{2}$$

Notice that we assume the data is generated by two steps:

- draw a latent variable x from the prior $p(x)$,
- draw data y from the likelihood $p(y|x)$.

3 Variational Inference

The **evidence** is the distribution over the whole data space, and is usually intractable to directly compute, **Variational Inference** is the technique to avoid calculating $p(y)$, instead, it assumes a model $f(\cdot)$ to simulate $p(x|y)$.