Prob/Stats Cheatsheet

Steve Young

Abstract: Everything I know about prob/stats/maybe information theory too..

Contents

1 Conventions

Math Notation

2 Distributions

2.1 Gaussians

1. To start with, *memorize* that

$$\boxed{\int_{-\infty}^{\infty} dx\, e^{-x^2} = \sqrt{\pi}} \tag{2.1}$$

2. Next, anything multiplying the x^2 in the integrand is present in inverse under the square root.

$$\int_{-\infty}^{\infty} dx\, e^{-\text{stuff}\, x^2} = \sqrt{\frac{\pi}{\text{stuff}}} \tag{2.2}$$

so, for example:

$$\int_{-\infty}^{\infty} dx\, e^{-\frac{1}{2}ax^2} = \sqrt{\frac{2\pi}{a}} \tag{2.3}$$

2.2 Bernoulli

For $x \in \{0, 1\}$, Bernoulli dist parametrized by μ, with

$$p(x; \mu) = \mu^x (1 - \mu)^{1-x} \tag{2.4}$$

3 Prob and stats

3.1 The Rules of Probability

- **Product Rule**: $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$
- **Sum Rule**: $p(x) = \sum_y p(x, y) = \sum_y p(x|y)p(y)$

3.2 Bayes' Rule

Using $p(y|x)p(x) = p(x, y) = p(x|y)p(y)$, we have

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_y p(x|y)p(y)} \tag{3.1}$$

3.3 Covariance

4 Information Theory

KL divergence:

$$\begin{aligned} KL\big[p(x)||q(x)\big] &= \sum_{x_i} p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) = -\sum_{x_i} p(x_i) \log\left(\frac{q(x_i)}{p(x_i)}\right) \\ &= -\sum_{x_i} p(x_i) \log q(x_i) + \sum_{x_i} p(x_i) \log p(x_i) \\ &= H(p, q) - H(p) \end{aligned} \tag{4.1}$$

where $H(p, q)$ is the cross entropy, and $H(p)$ is the entropy.

5 Bayesian

6 Optimal Stopping Theory