Prob/Stats Cheatsheet

Steve Young

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

Contents

1	Conventions	1
2	Distributions 2.1 Gaussians	
3	Prob and stats 3.1 The Rules of Probability	2
4	Information Theory	2
5	Bayesian	2
6	Optimal Stopping Theory	2

1 Conventions

Math Notation

2 Distributions

2.1 Gaussians

To start with, remember that

$$\int_{-\infty}^{\infty} dx \, e^{-\alpha x^2} = \left(\frac{\pi}{\alpha}\right)^{1/2}, \qquad \alpha = \frac{1}{2\sigma^2}$$
 (2.1)

where σ^2 is the variance of a Gaussian distribution.

2.2 Bernoulli

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

$$p(x;\mu) = \mu^{x} (1-\mu)^{1-x}$$
(2.2)

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

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

3.3 Covariance

4 Information Theory

KL divergence:

$$KL[p(x)||q(x)] = \sum_{x_i} p(x_i) \log \left(\frac{p(x_i)}{q(x_i)}\right) = -\sum_{x_i} p(x_i) \log \left(\frac{q(x_i)}{p(x_i)}\right)$$

$$= -\sum_{x_i} p(x_i) \log q(x_i) + \sum_{x_i} p(x_i) \log p(x_i)$$

$$= H(p, q) - H(p)$$

$$(4.1)$$

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

5 Bayesian

6 Optimal Stopping Theory