ML Background

Maximum likelihood etc.

September 4, 2017

Coin Tossing

- Given a coin, find out $P(heads)$
- I.e. the probability that if you flip it, it lands as 'heads'

Coin Tossing

- Given a coin, find out $P(heads)$
- I.e. the probability that if you flip it, it lands as 'heads'
- Flip it a few times: $H\ H\ T$
- $P(heads) = 2/3$, no need for Comp 379
- Hmm... is this rigorous?

Bernoulli distribution

- Single binary random variable $x \in \{0, 1\}$
- E.g. $x = 1$ represents 'heads' and $x = 0$ represents 'tails'
- Probability of $x = 1$ denoted by the parameter μ
- So, $p(x = 1|\mu) = \mu$ and $p(x = 0|\mu) = 1 - \mu$
- The probability distribution over x can be written

$$Bern(x|\mu) = \mu^x (1 - \mu)^{1-x}$$

Coin tossing model

- Assume coin flips are independent and identically distributed
- All are separate samples from the Bernoulli distribution (i.i.d.)
- Given data $\mathcal{D} = \{x_1, \dots, x_N\}$
- Where heads: $x_i = 1$ and tails: $x_i = 0$
- The **likelihood** of the data is:

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

Maximum Likelihood Estimation

- Given $\mathcal{D}$ with H heads and T tails
- What should μ be?
- Maximum Likelihood Estimation (MLE)
- Choose μ which maximizes the likelihood of the data

$$\mu_{ML} = \arg\max_{\mu} p(\mathcal{D}|\mu)$$

- Since $\ln(\cdot)$ is monotonically increasing:

$$\mu_{ML} = \arg\max_{\mu} \ln p(\mathcal{D}|\mu)$$

NOTE: A monotonically increasing function is one that increases as x does for all real x

Maximum Likelihood Estimation

- Likelihood

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

- Log-likelihood

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} x_n \ln \mu + (1-x_n)\ln(1-\mu)$$

- Take the derivative and set to 0

Maximum Likelihood Estimation

- Likelihood

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

- Log-likelihood

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} x_n \ln \mu + (1-x_n)\ln(1-\mu)$$

- Take the derivative and set to 0

$$\frac{d}{d\mu}\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} x_n \frac{1}{\mu} - (1-x_n)\frac{1}{1-\mu} = \frac{1}{\mu}H - \frac{1}{1-\mu}T$$

$$\mu = \frac{H}{T+H}$$

Acknowledgements: Slides based on the latex source provided by Oliver Schulte and Greg Mori (Simon Fraser University)