EE 126 — Probability and Random Processes

March 22, 2018 — MAP, MLE I

1 Lecture

Random Graphs

An E-R graph is specified as $G(n, p)$. n is the number of vertices in the graph. p is the edgewise probability of being connected, and is a function of n when it comes to the realization of certain properties.

The connectivity threshold is

$$p = \frac{\lambda \log n}{n}$$

When $\lambda > 1$, the graph should be connected and we should be able to go anywhere. In fact the probability of the graph being connected goes to 1. On the other hand, if $\lambda < 1$, the probability of the graph being unconnected (i.e. ≥ 1 isolated node) goes to 1.

To prove the first statement, we can use the idea that the graph being disconnected means "there exists a set of size k (where $1 \leq k \leq n/2$) s.t. there is no edge between the set and its complement." The strategy is then to enumerate all possible ways to be disconnected and overcount, and show that the overcount goes to 0.

$$\begin{aligned}
P(G(n,p) \text{ disconnected}) &= P\left(\bigcup_{k=1}^{n/2} (\exists \text{ a "smaller" set of size } k \text{ that is disconnected from its complement})\right) \\
&\leq \sum_{k=1}^{n/2} P(\exists \text{ a "smaller" set of size } k \text{ that is disconnected from its complement}) \\
&\leq \sum_{k=1}^{n/2} \binom{n}{k} P(\text{a \textit{specific set} of size } k \text{ is disconnected from the rest}) \\
&= \sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)} \to 0
\end{aligned}$$

We apply the union bound at several points in the above sequence, e.g. to omit the probabilities that multiple sets of size k are disconnected (i.e. the intersections).

Note: for more information, read the random graph notes. This is just a sketch.

Inference

There are two camps: *classical/frequentist* (where estimable quantities are values, fixed but unknown) and *Bayesian* (where estimable quantities are probabilistic).

Detection and Bayes' Rule

The basic premise of detection: there are N possible exclusive **causes** of a particular **symptom**. *Symptoms* are the observations; *causes* are what we blame the observations on. "Exclusive" means that there can only be one of the N causes for any given case.

Formally, each cause i has some prior probability p_i, a probability q_i of causing the observed symptom, and a posterior probability π_i given the observation (symptom) which is calculated using Bayes' rule:

$$\pi_i = P(C_i \mid S) = \frac{P(C_i)P(S \mid C_i)}{\sum_{j=1}^{N} P(S \mid C_j)P(C_j)} = \frac{p_i q_i}{\sum_{j=1}^{N} p_j q_j}$$

MAP / MLE

MAP refers to the *maximum a posteriori* estimate of cause given symptom. The MAP rule is $\arg\max_{i \in \{1,\ldots,N\}} p_i q_i$. The MLE (maximum likelihood estimation) rule is $\arg\max_i q_i$, which is the same as the MAP estimate if we assume a "uniform prior" ($p_i = \frac{1}{N}$ for each i).

More generally,

$$\text{MAP}[X \mid Y = y] = \arg\max_x P(X = x \mid Y = y)$$

i.e. "which cause best *explains* the observed symptom," while

$$\text{MLE}[X \mid Y = y] = \arg\max_x P(Y = y \mid X = x)$$

i.e. "which cause best *generates* the observed symptom." Y is the observation; X is the variable we want to estimate.

Example: bias of a coin. Bob tosses a coin three times and gets three heads. How should he estimate the bias of his coin? In other words, what is $P(H)$?

Suppose the coin is known to be either fair (i.e. $P(H) = \frac{1}{2}$) or two-headed (i.e. $P(H) = 1$).

$$q_F = P(3 \text{ heads} \mid \text{fair coin}) = \frac{1}{8}$$
$$q_B = P(3 \text{ heads} \mid \text{biased coin}) = 1$$

MLE says

$$\arg\max_{i \in \{F,B\}} q_i \implies \text{biased coin}$$

MAP says

$$\arg\max_{i \in \{F,B\}} p_i q_i = \arg\max \left\{ \left(\frac{9}{10}\right)\left(\frac{1}{8}\right), \left(\frac{1}{10}\right)(1) \right\} \implies \text{fair coin}$$

assuming priors $p_F = {}^9\!/_{10}$ and $p_B = {}^1\!/_{10}$.