Revision (Probabilistic AI)

The equations you need to remember

Logistic regression

Odds:

$$odds = \frac{p}{1-p}$$

Log-odds or logit:

$$\text{logit} = \log\left(\frac{p}{1-p}\right) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \cdots + \theta_n x_n$$

Prpbabolity:

$$p = \frac{1}{1+\frac{1}{odds}}$$

Information theory

Entropy:

$$\begin{aligned} H(X) &\equiv E\left[I_X(x)\right] \equiv -\sum_i^n P(X = x_i) \log_b P(X = x_i) \\ &\equiv E\left[\log_b \frac{1}{P_X(x)}\right] \equiv -E\left[\log_b P_X(x)\right] \end{aligned}$$

We use *b* = *2* in the exam.

Joint entropy:

$$H(X,Y) = -E[\log p(X,Y)] = -\sum_{x_i \in R_X} \sum_{y_j \in R_Y} p(x_i, y_j) \log p(x_i, y_j),$$

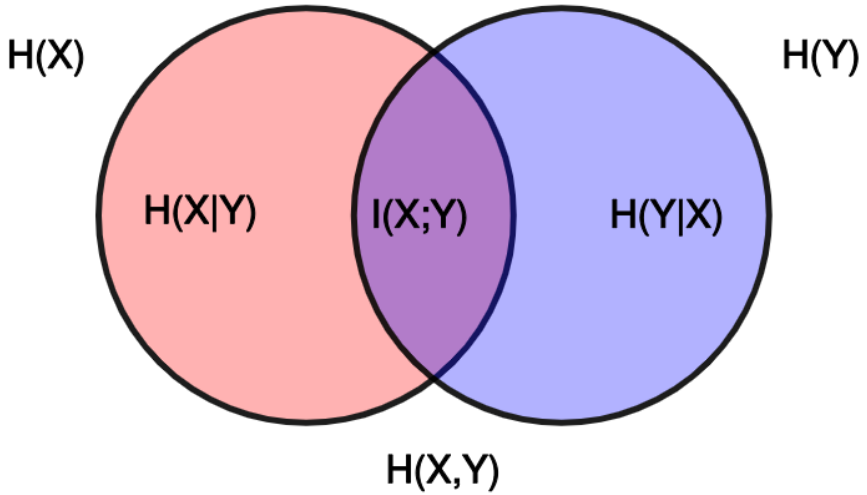
Conditional entropy:

$$H(Y|X) = H(X,Y) - H(X)$$
$$H(X|Y) = H(X,Y) - H(Y)$$

Mutual information:

$$\begin{aligned} I(X;Y) &\equiv H(X) - H(X|Y) \\ &\equiv H(Y) - H(Y|X) \\ &\equiv H(X) + H(Y) - H(Y,X) \\ &\equiv H(X,Y) - H(Y|X) - H(X|Y) \end{aligned}$$

This Venn diagram is useful



Bayes Theorem

$$p(\Theta|x) = \frac{p(x|\Theta)p(\Theta)}{p(x)}$$

Bayesian networks:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\mathbf{Parents}(X_i))$$

Topics

Here are the critical bits in probabilistic AI for your to prepare for your exam:

Week 1 Lecture 2:

1. Concepts of likelihood， page 5

2. Maximum likelihood estimator, the meaning of equation 5, page 9

3. Cost function of Maximum likelihood estimator, page 11

4. Why we use the negative logarithm of the likelihood function, page 12

Week 7 Lecture 1

1. Concepts of odds and logit, page 7

2. Interpretation of fitted Logistic regression model，17-19

3. Main assumptions of logistic regression, page 20

Week 7 Lecture 2: Introduction to Information Theory

1. Concept of self-information, page 9
2. Concept of entropy, page 11

Week 8 Lecture 1: Information Theory II -- Measures for more variables

1. Joint entropy, page 4
2. Chain rule for conditional entropy, page 8
3. Mutual information and entropy, pages 15-16
4. Exercise question: 18 (**model answers**)

Week 8 Lecture 2: Information Theory II -- Applications of Information Theory to Machine Learning

1. Information gain is mutual information, page 19
2. Mutual information feature selection, page 21

Week 9 Lecture 1: Bayes theorem revisit

1. Example, pages 6-8, 10, 12, 16.

Week 9 Lecture 2: Bayesian networks

1. Bayesian network (BN) representation, page 8
2. Wet grass example, page 9
3. Bayesian network joint probability equation, page 10, equation 2.
4. Casual structures, pages 11, 12-13, 17-18, 20-23