Chapter 1.8-1.10

Sagar Malhotra

April 4, 2020

Some useful results from probability theory

We will discuss a bit of naotation.

- **Probability Density:** $P(u, v)$
- **Marginal Probability Density:** $P(u) = \int p(u, w)dw$
- **Factoring Joint PD:** $P(u, v, w) = p(u|v, w)p(v, w) = p(u|v, w)p(v|w)p(w)$
- Every **PD** in BDA is conditioned on a certain Hypothesis, **H**
- In BDA we talk about the distribution of the target variable and the model parameters for example $p(y, \theta)$. And $p(y)$ is different from $p(y|\theta)$
- But even more importantly we hide the conditioning on the hypothesis.

$$p(\theta, y|\mathbf{H}) = p(\theta|\mathbf{H})p(y|\theta, \mathbf{H})$$

- Here **H** refers to the set of Hypothesis or assumptions. Please refer to Bishop page 26 to page 31 to see this in action.

$$E(u) = \int up(u)du$$

$$\text{var}(u) = \int (u - E(u))^2 p(u)du$$

- Covariance Matrix

$$\text{var(u)} = \int (u - E(u))(u - E(u))^T p(u)du$$

- Note that u is a $d \times 1$ column vector and $d \times 1$ vector times $(d \times 1)^T$ is a $d \times d$ matrix.

Modelling using conditional probability

- Importance of Hierarchical thinking.
- Example of Hierarchical modelling in students height distribution

$$p(height, gender) = p(height|geneder)p(gender)$$

Means and variances of conditional distributions

How to express mean and variance of a given variable with respect to its conditional distribution with respect to some given variable.

$$E(u) = E(E(u|v))$$

Deriving this expression.

$$E[u] = \int\int up(u,v)dudv = \int\int up(u|v) \quad du \quad p(v)dv$$

Now,

$$E[u|v] = \int up(u|v)du$$

Hence,

$$E[u] = \int E[u|v]p(v)dv$$

Variance

The case for variance has two terms

$$\mathrm{var}(u) = E[\mathrm{var}(u|v)] + \mathrm{var}(E[u|v])$$

In words: the variance of u is the sum of the expected conditional variance u given v and the variance of the conditional expectation of v given u. The first term captures the variation left after "using v to predict u", while the second term captures the variation due to the mean of the prediction of u due to the randomness of v.

Jacobian: Transformation of Variables

It is common to transform the probability spaces in order to get a more comfertable view of the

Sampling using the inverse cumulative distribution function

Let's say I have a uniform random number generator and I want to generate a set of random number distributed according to given probability distribution, How can we do that? Pause and think and realise that its not so trivial !!