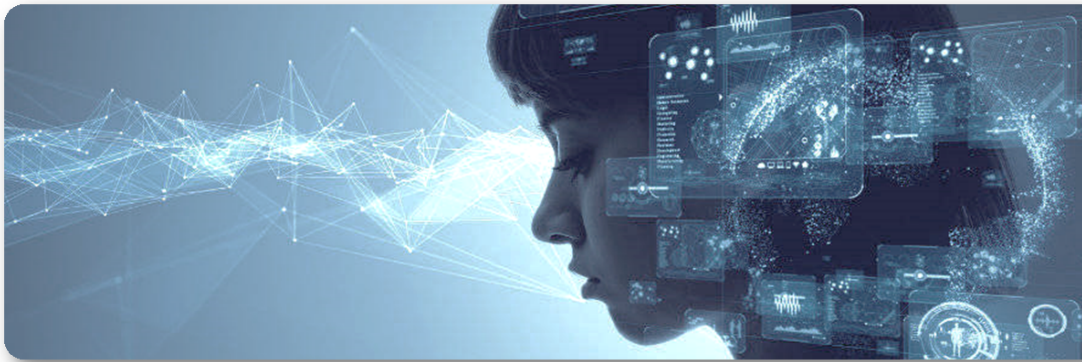

COMP 2211 Exploring Artificial Intelligence

Supplementary Notes:

Naïve Bayes Classifier - Why $P(e_i|B_i)$ can be found by substituting $x = e_i$ to $f(x)$?

Prof. Song Guo, Dr. Desmond Tsoi & Dr. Huiru Xiao

Department of Computer Science & Engineering
HKUST, Hong Kong SAR, China

Explanation

- In continuous probabilities, the probability of getting precisely any given outcome is 0, and this is why densities are used instead.
- Therefore, we do not deal with expressions such as $P(X = x)$ but with $P(|X - x| < \Delta(x))$, which stands for the probability of X being close to x.
- Let's simplify the notation and write $P(X \sim x)$ for $P(|X - x| < \Delta(x))$.
- If we apply the Bayes' rule here, we will get

$$P(X \sim x|W \sim w) = \frac{P(W \sim w|X \sim x)P(X \sim x)}{P(W \sim w)}$$

- Because we are dealing with probabilities. If we now introduce densities:

$$pdf(x|w)\Delta(x) = \frac{pdf(w|x)\Delta(w)pdf(x)\Delta(x)}{pdf(w)\Delta(w)}$$

Explanation

- Since probability $=$ density $\times$ neighborhood-size
- Also, since all $\Delta(\cdot)$ cancel out in the expression above, we get

$$pdf(x|w) = \frac{pdf(w|x)pdf(x)}{pdf(w)}$$

which is the Bayes rule for densities.

Explanation

- Now, it comes down to integral approximations.
- To get the probability of a specific variable value from the variable's continuous probability density function (PDF), we integrate the PDF around the value in question over an interval of width epsilon, and take the limit of that integral as epsilon approaches 0.
- For small epsilon, this integral will be equivalent to the product of epsilon and the height of the PDF at the variable value in question.
- Ordinarily, the limit of this expression would be to 0 as epsilon approached 0. However, in the case of Naive Bayes we are interested in the ratio of conditional probabilities.
- Because both the numerator and denominator of our ratio will include a factor of epsilon, these factors of epsilon cancel out.
- As a result, the limit of the ratio of conditional probabilities will be equivalent to the ratio of the PDF heights at the variable value in question.

For more rigorous explanation

`https://github.com/ashkonf/CompleteNB/blob/master/Derivation.pdf`

Explanation

- The conclusion is that, given that the Bayes rule also holds for densities, it is legitimate to use the same methods replacing probabilities with densities when dealing with continuous random variables.

This set of lecture notes is written based on various resources in the Internet.

That's all!

Any questions?