N. Kausik
CS21M037

# Tutorial 2

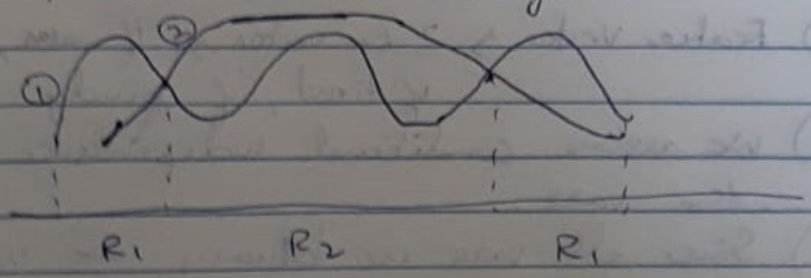
1. In discriminative model → we find Conditional Prob $P(y/x)$
   In generative model → we find Joint prob $P(y, x)$

2. Prior prob = $\frac{\#\text{ occurrence of particular class}}{\text{Total } \#\text{ samples}}$

3. Yes, evidence can be ignored as we divide by same value for all classes and we only consider prediction by max of classes.

4. Naive Bayes chooses the class with max conditional prob and hence it is optimal

5. No, need not be contiguous
   Eg:




6. $R_1 - C_1$ $R_4 - C_2$
   $R_2 - C_2$ $R_5 - C_1$
   $R_3 - C_1$ $R_6 - C_2$

7. (Loss) for class,
   $S_1 \Rightarrow 0.1\times0 + 0.07\times1 + 0.04\times2 + 0.01\times3$
   $= 0.07 + 0.08 + 0.03 = 0.18$
   Similarly,
   $S_2 \Rightarrow 10.06 = 0.1\times100 + 0.07\times0 + 0.04\times1 + 0.01\times2$
   $S_3 \Rightarrow 27.01 = 0.1\times200 + 0.07\times100 + 0.04\times0 + 0.01\times1$
   $S_4 \Rightarrow 48 = 0.1\times300 + 0.07\times200 + 0.04\times100 + 0.01\times0$

8. a) Feature Vector ⇒ Frequencies of words in the mail

b) We assume conditional independence b/w all the words as we are using freq of words.

c) Multinomial distribution as we use freq of words which is a discrete dist.

9. a) Feature Vector ⇒ 5 dim vector of amount of sugar, baking powder, oil, temp and duration

b) We assume conditional independence b/w the 5 values in vector.

c) Since amount of sugar, etc are continuous variables, we use Gaussian Distribution.

10. a) Feature Vector ⇒ 3 dim vector of if warm, if raining, if tired, if windy

b) We assume conditional independence b/w the vars.

c) Since all vars are binary, we use Bernoulli dist.

11. Not practical for large number of data points as for every point we need to check its distance with every other point to find k-nearest points.

It will take lot of time for more points.