## How Naive Bayes handles numerical data?

Ex.

| age | married |
|---|---|
| 27 | Y |
| 61 | N |
| 52 | Y |
| ⋮ | ⋮ |

- New Query pt ⇒ 55, tell if a person is married or not.
- In this case also 2 prob will be calculated

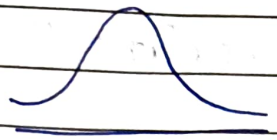
⇒ $P(Y|55)$ & $P(N|55)$

⇒ $P(Y|55) = \underbrace{P(55|Y)}\, P(Y)$

= In this the challenge is this

- In categories, it was present, but in case of numbers, you can't find 55 hence how $P(55|Y)$ will be calculated
- In numerical, we assume that the particular age column (numerical)

will follow Gaussian Distribution




Since we are finding P(55|Y), mean & std of 'Yes' class will be calculated

- We will find it's $\mu$, $\sigma$
- Since we assumed it to be Gaussian, we have its PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

this is probability density

- We have $\mu$, $\sigma$, and $x = 55$
- for 55 we will get some Probability density

$$f(55) = 0.62 - = P(55|Y)$$

- We have $\rightarrow P(Y|55) = P(55|Y) \times P(Y)$

$$P(Y|55) = 0.62 \times P(Y) \quad \text{(1)}$$

- similarly we will find for P(55|N).

P(N|55)

$$\Rightarrow P(N|55) = P(55|N) \times P(N)$$

- put mean & std ~~and X~~ of NO class in Gaussian pdf.

$$P(55|N) = f(55) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{55-\mu}{\sigma}\right)^2}$$

let's say $\Rightarrow$ 0.20

$P(55|N) = 0.2$

$P(N|55) = 0.2 \times P(N)$ — (2)

- whosever Prob is higher will be assigned that probability

* Now, what if age column is actually not Gaussian?

a) Data transformations: good if one can change distribution to Gaussian Dist

b) Alternative Distributions:
- Try out Different Distributions PDF such as exponential, poisson

c) Binning / Discretization

convert your numerical col to categorical, which solves the problem of Gaussian Distribution

d) KDE:
- understand the distribution of column By understanding if you can find Eqn for which you can calculate probability densities just like PDF of distributions

e) Use other models