Naïve Bayes

Bayes Theorem

- **Given a hypothesis *H* and evidence *E* for this hypothesis, then the probability of *H* given *E*, is:**

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

Example: Given are instances of fruits, described by their color and shape. Let:

- *E* is red and round

- *H* is the hypothesis that *E* is an apple

Bayes Theorem

Example: Given are instances of fruits, described by their color and shape. Let:

- *E* is red and round

- *H* is the hypothesis that *E* is an apple

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

- ***P(H|E)*** **is the probability that *E* is an apple, given that we have seen that *E* is red and round**
  - **Called *posterior probability* of *H* conditioned on *E***

- ***P(H)*** **is the probability that any given example is an apple, regardless of how it looks**
  - **Called *prior probability* of *H***

Exercise 1

Homework

P1	P2	Result
Y	Y	good
Y	N	bad
N	N	good
N	Y	bad
Y	N	good
N	N	good

New item to classify:

P1=N, P2=Y

E1 is P1=N, E2 is P2=Y

We need to compute

P(good|E) and **P(bad|E)** and compare them.

E1 is P1=N, E2 is P2=Y

Compute **P(good|E)**

P1	P2	Result
Y	Y	good
Y	N	bad
N	N	good
N	Y	bad
Y	N	good
N	N	good

$$P(good \mid E) = \frac{P(E_1 \mid good)\, P(E_2 \mid good)\, P(good)}{P(E)}$$

E1 is P1=N, E2 is P2=Y

Compute **P(bad|E)**

P1	P2	Result
Y	Y	good
Y	N	bad
N	N	good
N	Y	bad
Y	N	good
N	N	good

$$P(good \mid E) = \frac{P(E_1 \mid good)\,P(E_2 \mid good)\,P(good)}{P(E)}$$

P(good)=4/6=2/3
P(E1|good)=P(P1=N|good)=2/4=1/2
P(E2|good)=P(P2=Y|good)=1/4

P(bad)=2/6=1/3
P(E1|bad)=P(P1=N|bad)=1/2
P(E2|bad)=P(P2=Y|bad)=1/2

$$P(good \mid E) = \frac{\frac{1}{2}\frac{1}{4}\frac{2}{3}}{P(E)} = \frac{\frac{1}{12}}{P(E)}$$

$$P(bad \mid E) = \frac{\frac{1}{2}\frac{1}{2}\frac{1}{3}}{P(E)} = \frac{\frac{1}{12}}{P(E)}$$

The two probabilities are the same. To resolve the tie we randomly choose between the 2 classes => e.g. class *good*.

Exercise 2

Exercise 2

"Naïve" Definition (Cambridge Dictionary)

"too willing to believe that someone is telling the truth, that people's intentions in general are good, or that life is simple and fair…"

Why is Naïve Bayesian classification called "naïve"?

Exercise 2

<u>“Naïve” Definition</u> <u>(Cambridge Dictionary)</u>

“too willing to believe that someone is telling the truth, that people's intentions in general are good, or that life is simple and fair…”

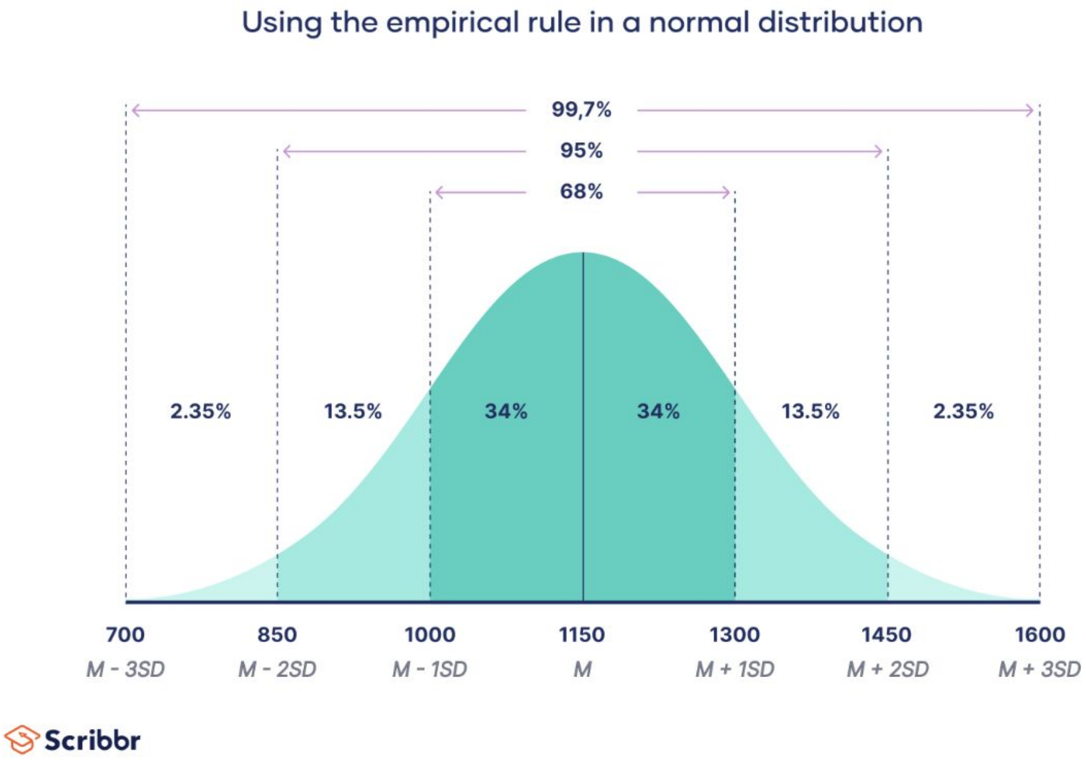
Why is Naïve Bayesian classification called “naïve”?

Naïve Bayes assumes that the values of the attributes are **independent** of each other and that all attributes are **equally important**. These assumptions are <u>unrealistic</u>, that’s why it is called “Naïve”.

Exercise 3

NB with numerical & nominal attributes

Gaussian Distribution



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

Mean - the middle point of the data
S.D. - reflect the spread of the data from the middle point

outlook=overcast, temperature=60, humidity=62, windy=false

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
overcast	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

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

σ (sigma): standard deviation
μ (mu): mean

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n-1}}$$

outlook=overcast, temperature=60, humidity=62, windy=false

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
overcast	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

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

μ_temp_yes = 73
σ_temp_yes = 6.2

f(temp=60|yes) = ?

outlook=overcast, temperature=60, humidity=62, windy=false

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
overcast	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

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

μ_temp_yes = 73
σ_temp_yes = 6.2

f(temp=60|yes)

$$= \frac{1}{6.2\sqrt{2\pi}} e^{-\frac{(60-73)^2}{2*6.2^2}} = 0.0071$$

outlook=overcast, temperature=60, humidity=62, windy=false

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

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
overcast	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

μ_temp_yes = 73
σ_temp_yes = 6.2

μ_hum_yes = 79.1
σ_hum_yes = 10.2

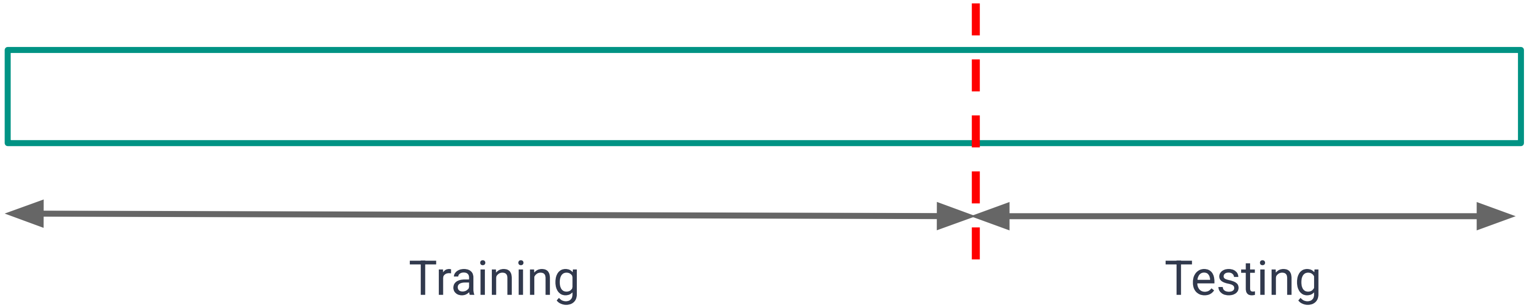
μ_temp_no = 74.6
σ_temp_no = 8.0

μ_hum_no = 86.2
σ_hum_no = 9.7

See Official Solution for Exercise 3

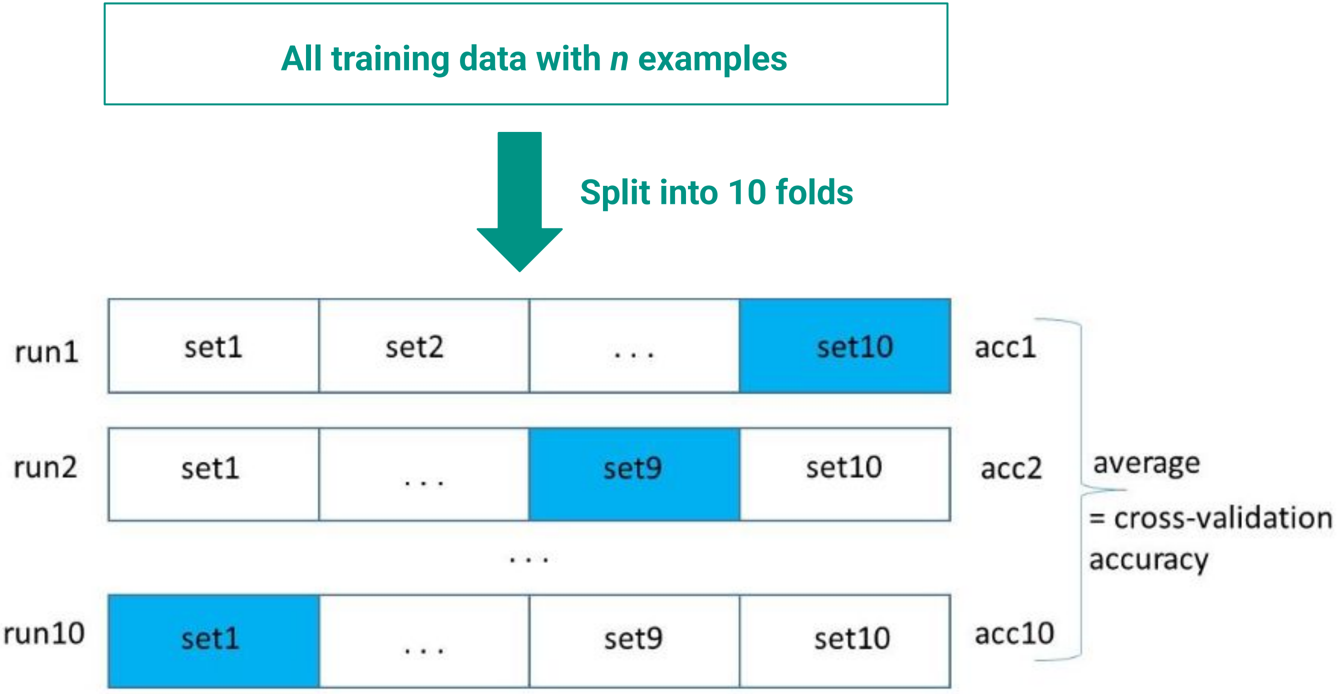
Evaluation

Single Train Test Split



Weka default: 66-34
In research: 75-25, 80-20, etc.

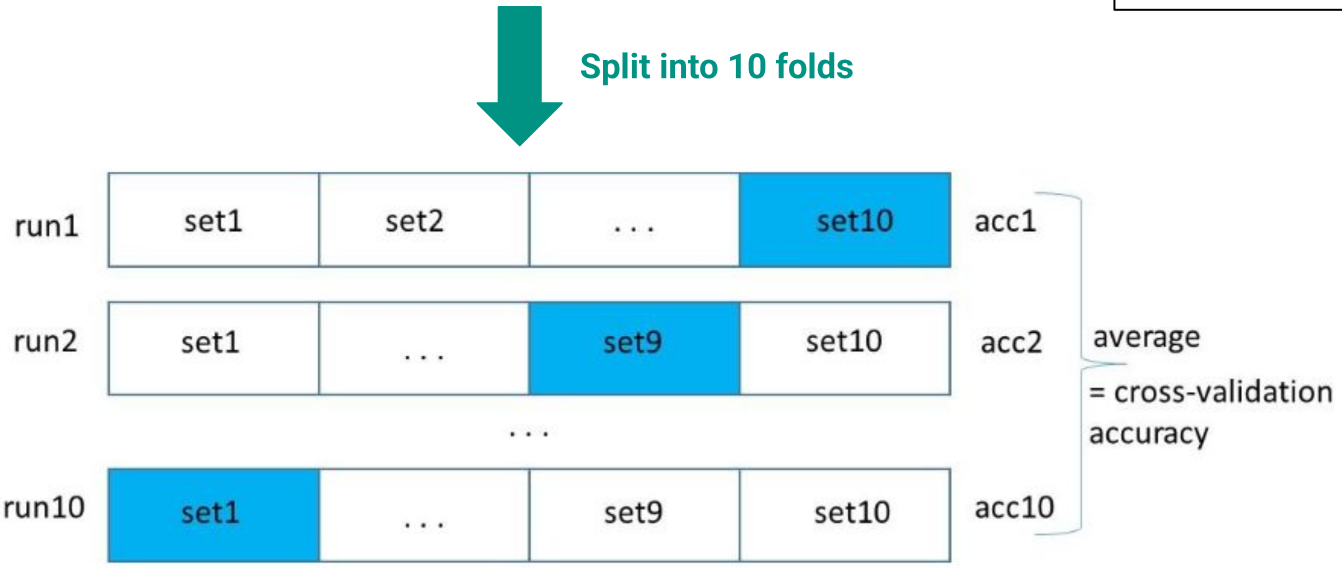
10-fold Cross validation



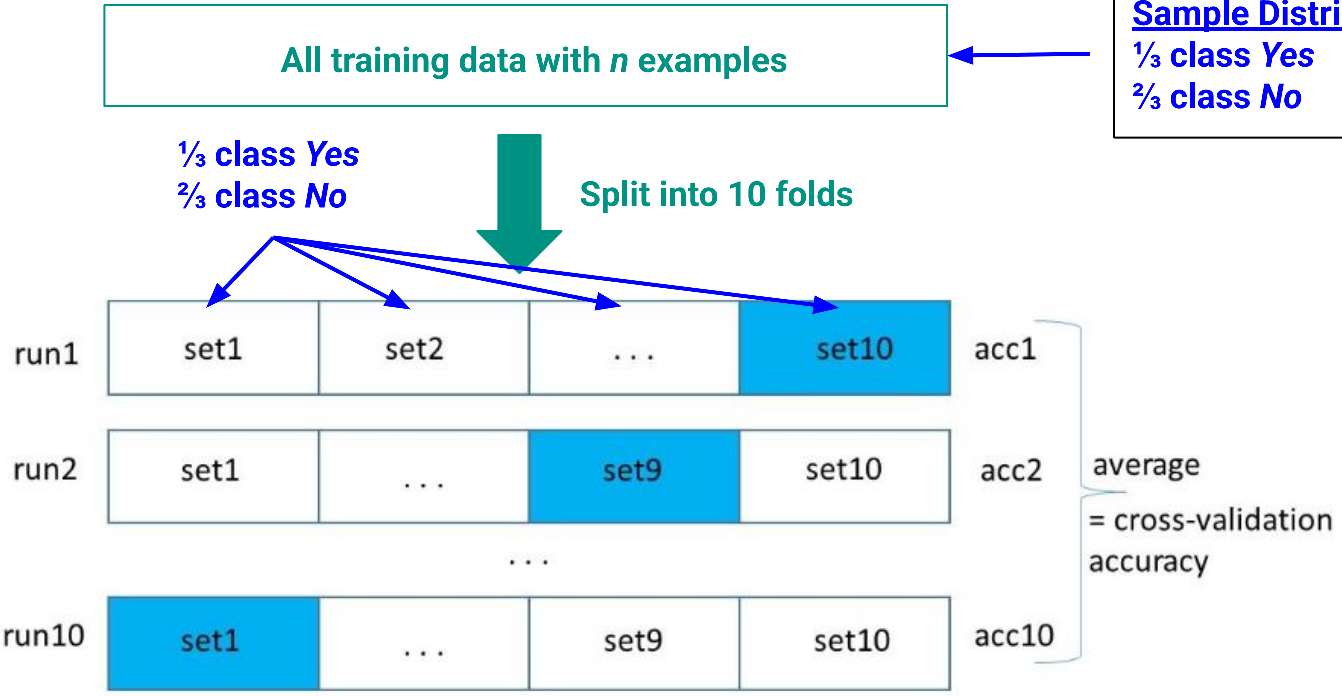
Stratified 10-fold Cross validation – Example

All training data with *n* examples

Sample Distribution:
⅓ class *Yes*
⅔ class *No*



Stratified 10-fold Cross validation - Example



Sample Distribution:
⅓ class *Yes*
⅔ class *No*

Leave-one-out Cross validation