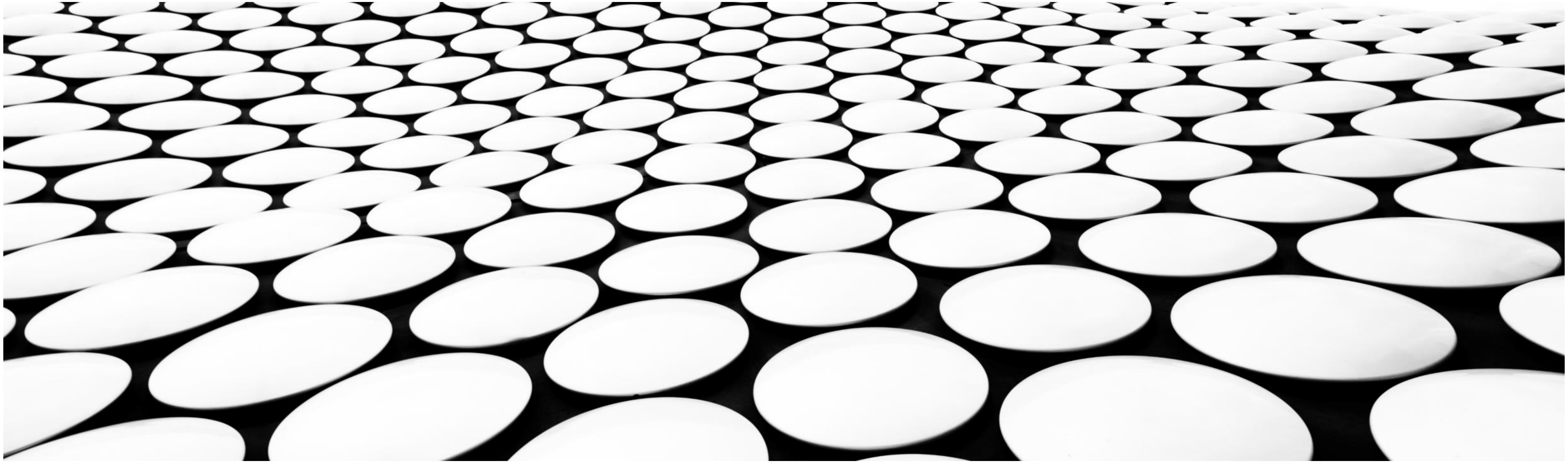
NAIVE BAYES CLASSIFIER

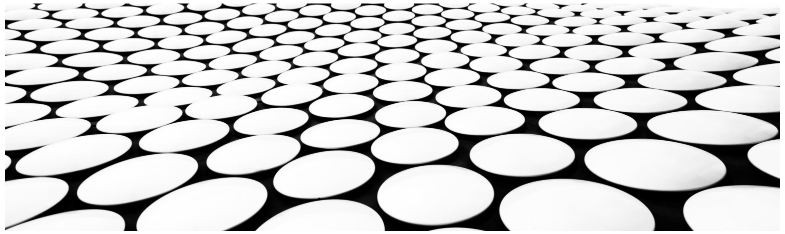
PART 3 OF 3 – CONDITIONAL INDEPENDENCE



GOALS

- Bike/Drive Example
- Conditional independence

BIKE/DRIVE EXAMPLE

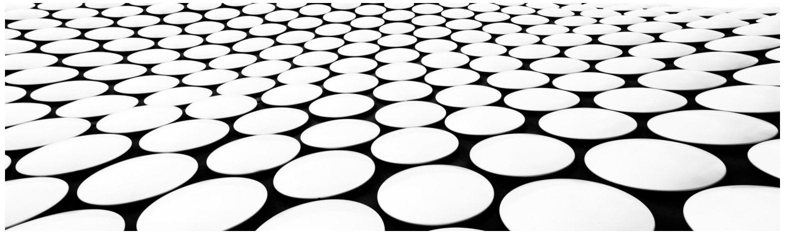


<u>Temperature:</u>
Very Cold (-30,-10)
Cold (-10, 0)
Average (0, 10)
Warm (10,20)
Very Warm (20, 35)

<u>Rain:</u>
Little (0,5)
Normal (5,20)
Lot (20,50)

Sample	Temperature	Rain/snow	Bike / Drive
s1	Very Cold	Normal	Drive
s2	Very Cold	Normal	Drive
s3	Very Cold	Normal	Drive
s4	Cold	Normal	Drive
s5	Cold	Lot	Drive
s6	Average	Lot	Drive
s7	Cold	Little	Bike
s8	Average	Little	Bike
s9	Warm	Little	Bike
s10	Average	Little	Bike
s11	Warm	Normal	Bike
s12	Warm	Normal	Bike
s13	Warm	Lot	Bike
s14	Warm	Little	Bike
s15	Warm	Normal	Bike

BIKE/DRIVE EXAMPLE



Two hypotheses: Drive or Bike

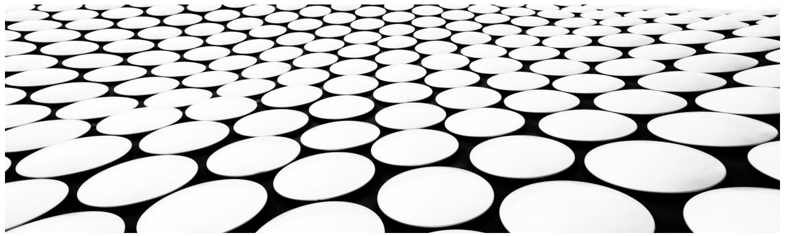
p(Bike|(E) = (p(E|Bike)* p(Bike)) / p(E)

Max

p(Drive|(E) = (p(E|Drive)* p(Drive)) / p(E)

Sample	Temperature	Rain/snow	Bike / Drive
s1	Very Cold	Normal	Drive
s2	Very Cold	Normal	Drive
s3	Very Cold	Normal	Drive
s4	Cold	Normal	Drive
s5	Cold	Lot	Drive
s6	Average	Lot	Drive
s7	Cold	Little	Bike
s8	Average	Little	Bike
s9	Warm	Little	Bike
s10	Average	Little	Bike
s11	Warm	Normal	Bike
s12	Warm	Normal	Bike
s13	Warm	Lot	Bike
s14	Warm	Little	Bike
s15	Warm	Normal	Bike

NAIVE BAYES CLASSIFIER



Since the denominator p(E) is the same for all classes:

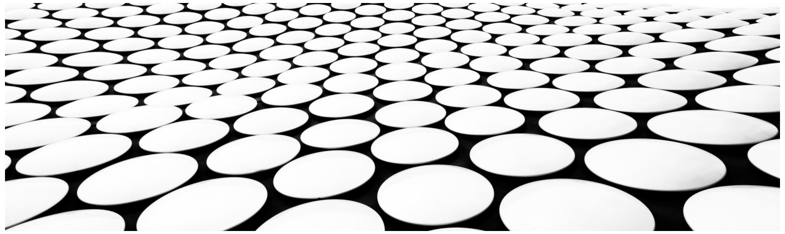
$$p(h_i|E) = \frac{p(E|h_i) \times p(h_i)}{\sum_{k=1}^{n} p(E|h_k) \times p(h_k)}$$

We only calculate the numerator: $\mathbf{p(h_i \mid E) \propto p(E \mid h_i) \times p(h_i)}$ *(we replace = by $\propto$)*

We test for
argmax(h_i) of $p(E \mid h_i) \times p(h_i)$

Read argmax as "the index i corresponding to the maximum value of h_i".

EXAMPLE OF CLASSIFICATION



Two hypotheses: Drive or Bike

$p(\text{Bike}|(E) \propto p(E|\text{Bike})* p(\text{Bike})$

Max

$p(\text{Drive}|(E) \propto p(E|\text{Drive})* p(\text{Drive})$

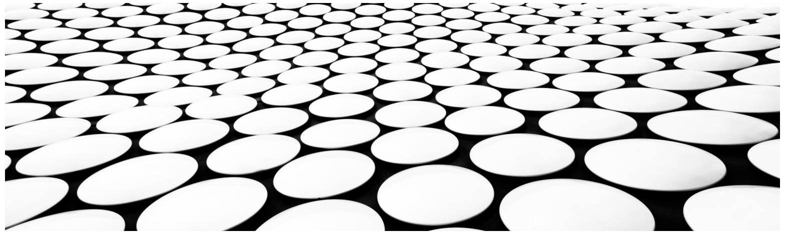
Suppose that the event E is (T=Average)

$$\boxed{\begin{aligned} P(\text{Bike}|(\text{Average}) &\propto P(\text{Average}|\text{Bike}) * P(\text{Bike}) \\ &\propto 2/9 * 9/15 \\ &\propto 0.13 \end{aligned}}$$

$$\begin{aligned} P(\text{Drive}|(\text{Average}) &\propto P(\text{Average}|\text{Drive})\ P(\text{Drive}) \\ &\propto 1/6 * 6/15 \\ &\propto 0.06 \end{aligned}$$

Sample	Temperature	Rain/snow	Bike / Drive
s1	Very Cold	Normal	Drive
s2	Very Cold	Normal	Drive
s3	Very Cold	Normal	Drive
s4	Cold	Normal	Drive
s5	Cold	Lot	Drive
s6	Average	Lot	Drive
s7	Cold	Little	Bike
s8	Average	Little	Bike
s9	Warm	Little	Bike
s10	Average	Little	Bike
s11	Warm	Normal	Bike
s12	Warm	Normal	Bike
s13	Warm	Lot	Bike
s14	Warm	Little	Bike
s15	Warm	Normal	Bike

CONDITIONAL INDEPENDENCE ASSUMPTION



Two hypotheses: Drive or Bike

$p(\text{Bike}|(E) \propto p(E|\text{Bike}) * p(\text{Bike})$

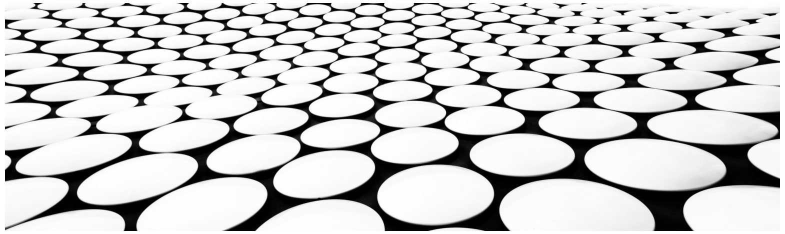
Max

$p(\text{Drive}|(E) \propto p(E|\text{Drive}) * p(\text{Drive})$

Suppose that event E is (T=Average AND R=Lot)

Sample	Temperature	Rain/snow	Bike / Drive
s1	Very Cold	Normal	Drive
s2	Very Cold	Normal	Drive
s3	Very Cold	Normal	Drive
s4	Cold	Normal	Drive
s5	Cold	Lot	Drive
s6	Average	Lot	Drive
s7	Cold	Little	Bike
s8	Average	Little	Bike
s9	Warm	Little	Bike
s10	Average	Little	Bike
s11	Warm	Normal	Bike
s12	Warm	Normal	Bike
s13	Warm	Lot	Bike
s14	Warm	Little	Bike
s15	Warm	Normal	Bike

CONDITIONAL INDEPENDENCE ASSUMPTION



Independents Events

Two events A and B are independent if and only if the probability of their joint occurrence is equal to the product of their individual (separate) occurrence.

General form: $p(A \cap B) = P(A|B) * P(B)$

Independence of A and B: $p(A \cap B) = P(A) * P(B)$

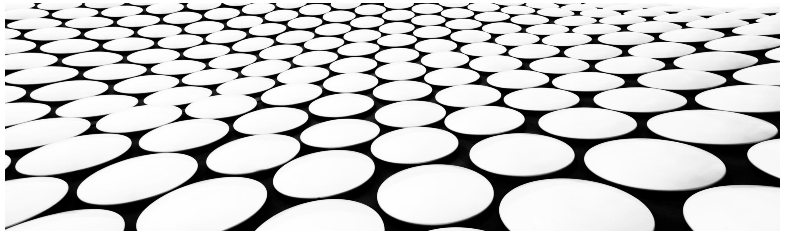
When $P(B) \neq 0$, it is the same as saying $P(A) = P(A|B)$

Conditionally Independents Events

Two events A and B are conditionally independent of each other given a third event C if and only if:

$p((A \cap B)|C) = p(A|C) * p(B|C)$

CONDITIONAL INDEPENDENCE ASSUMPTION



Sample	Temperature	Rain/snow	Bike / Drive
s1	Very Cold	Normal	Drive
s2	Very Cold	Normal	Drive
s3	Very Cold	Normal	Drive
s4	Cold	Normal	Drive
s5	Cold	Lot	Drive
s6	Average	Lot	Drive
s7	Cold	Little	Bike
s8	Average	Little	Bike
s9	Warm	Little	Bike
s10	Average	Little	Bike
s11	Warm	Normal	Bike
s12	Warm	Normal	Bike
s13	Warm	Lot	Bike
s14	Warm	Little	Bike
s15	Warm	Normal	Bike

Two hypotheses: Drive or Bike

$p(Bike|(E) \propto p(E|Bike)^* \, p(Bike)$

Max

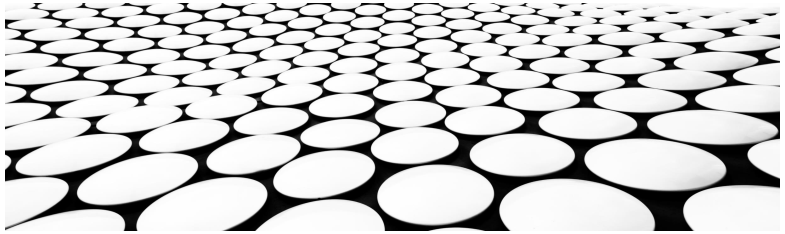
$p(Drive|(E) \propto p(E|Drive)^* \, p(Drive)$

E = (T=Average AND R=Lot)

$$
\begin{aligned}
P(Bike|(Average \,\&\, Lot) &\propto P(Average \,\&\, Lot|Bike)\, P(Bike) \\
&\propto P(Average|Bike) * P(Lot|Bike) * P(Bike) \\
&\propto 2/9 * 1/9 * 9/15 \\
&\propto 0.015
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
P(Drive|(Average \,\&\, Lot) &\propto P(Average \,\&\, Lot|Drive)\, P(Drive) \\
&\propto P(Average|Drive) * P(Lot|Drive) * P(Drive) \\
&\propto 1/6 * 2/6 * 6/15 \\
&\propto 0.022
\end{aligned}
}
$$

IN SUMMARY



	Naive Bayes
Underlying theory	Probability theory
Discrete or Continuous features ?	Discrete
Multi-value classes?	Yes
Learning process	Calculate the prior (hypothesis) and posterior (features|hypothesis)
Reasoning process	Apply Bayes Theorem, calculate posterior on hypotheses
Sensitive to sampling technique	YES. As a probabilistic classifier, the priors play a large role in classification.

IN SUMMARY

- Naive Bayes Classifier
  - Review of probabilities (part 1)
  - Bayes theorem and Naive Bayes Classifier (part 2)
  - Conditional independence (part 3)