NAÏVE BAYES MODELS

Dr. Aric LaBarr

Institute for Advanced Analytics

GENERAL IDEA

Naïve Bayes Classification

- When we need to classify observations there are two different sources of evidence:
  1. Similarity to other observations based on certain metrics/variables.
  2. Past decisions on classifications of observations like it.

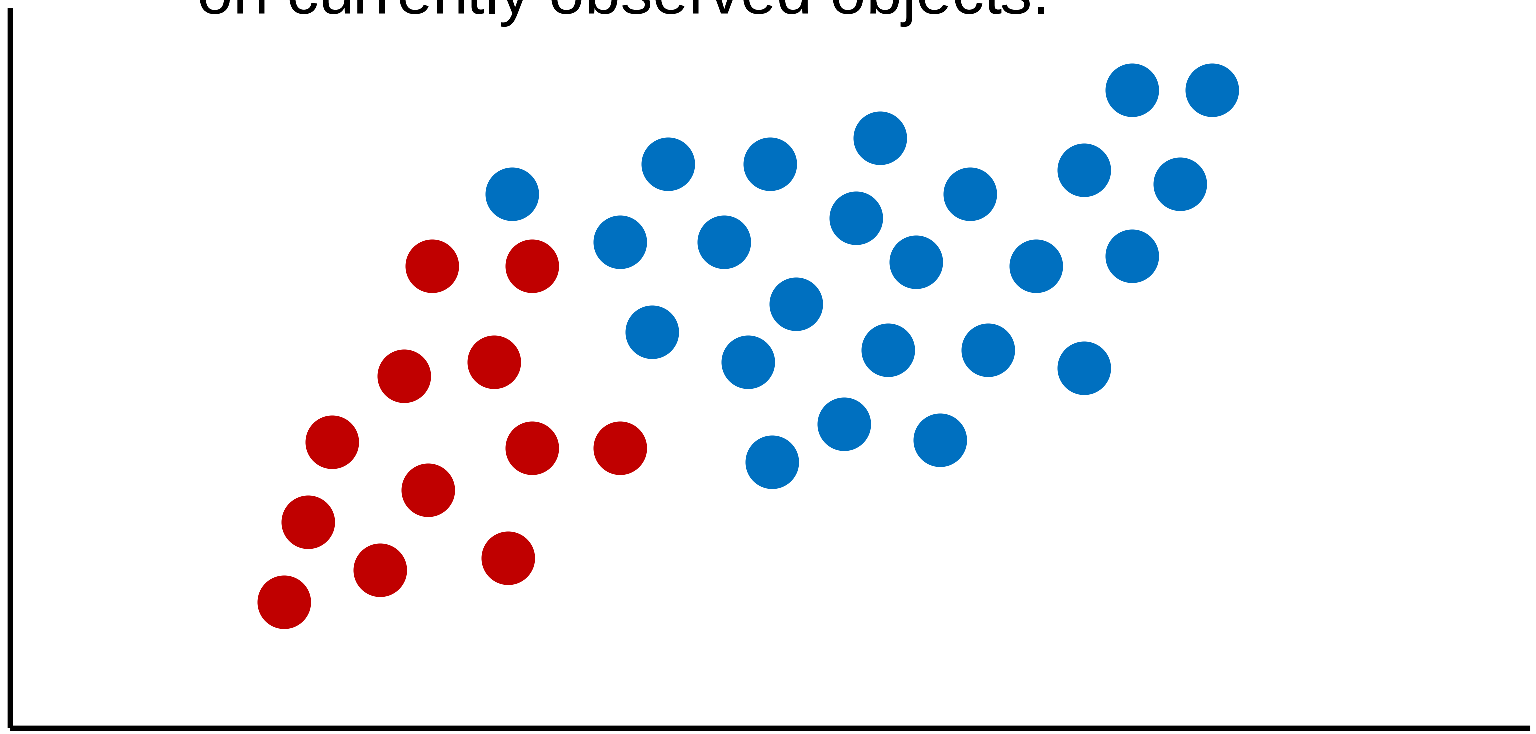
Naïve Bayes Classification

- When we need to classify observations there are two different sources of evidence:
  1. Similarity to other observations based on certain metrics/variables.
  2. Past decisions on classifications of observations like it.

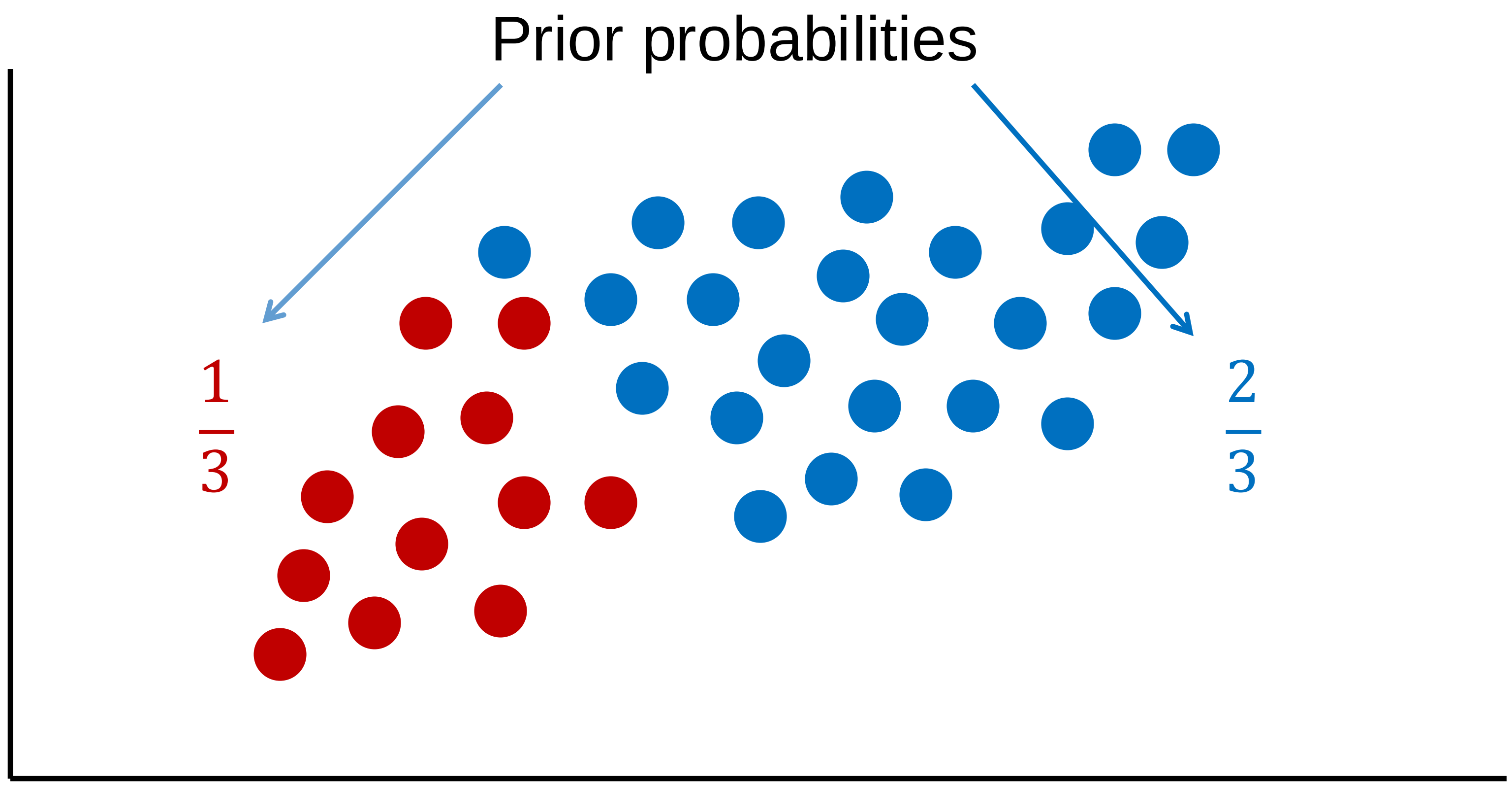
Bayesian approach, not frequentist.

Naïve Bayes Classification

Want to classify new observations based on currently observed objects.

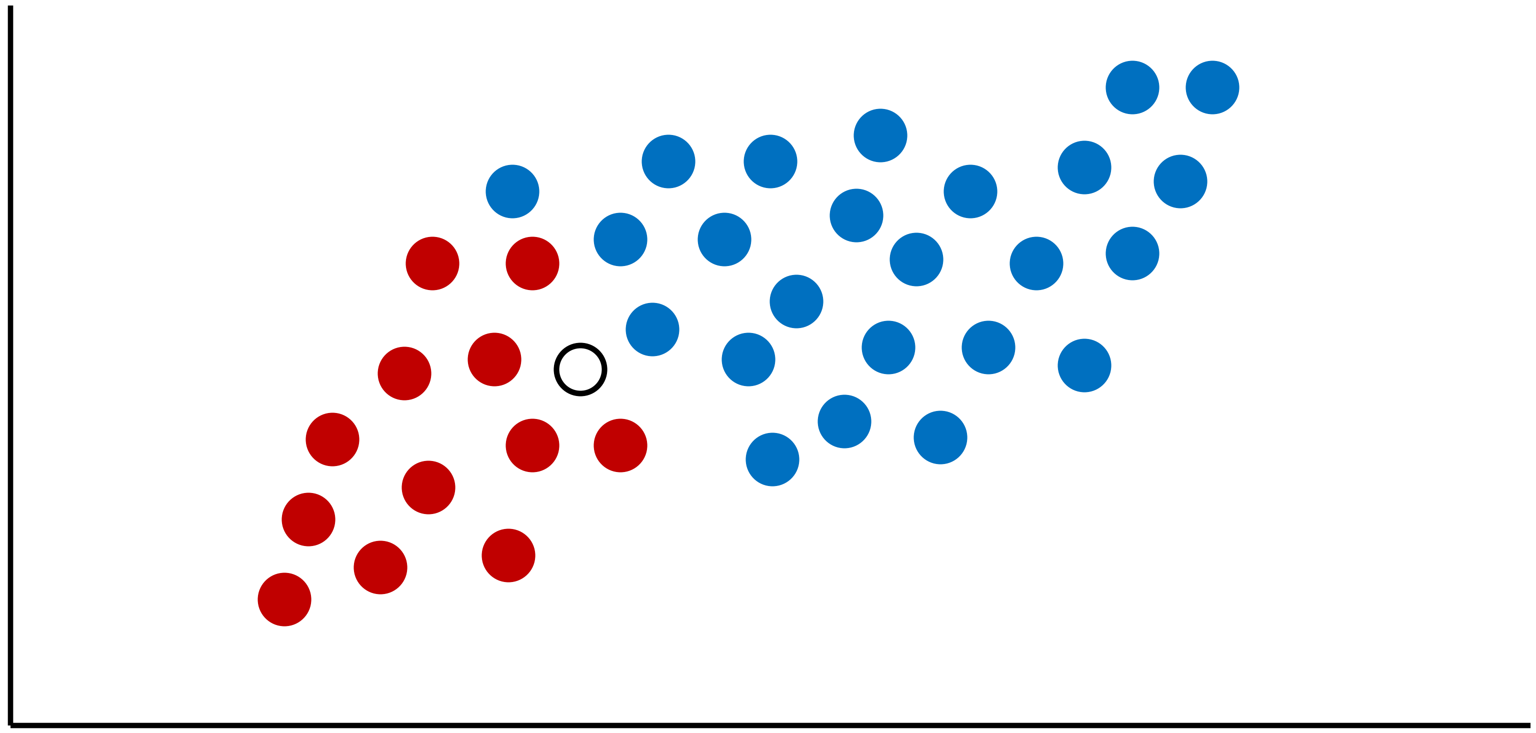


Naïve Bayes Classification



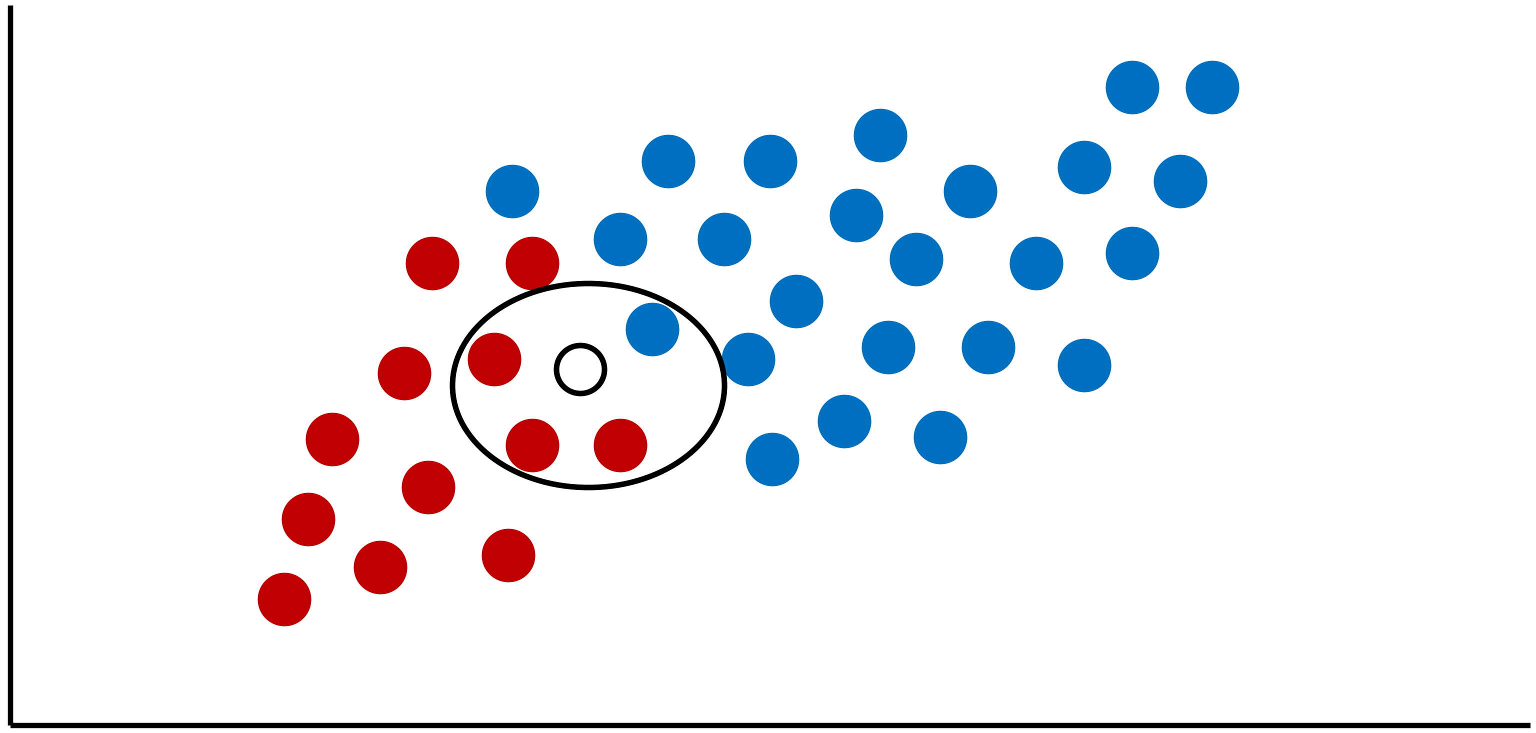
Naïve Bayes Classification

New observation to classify.

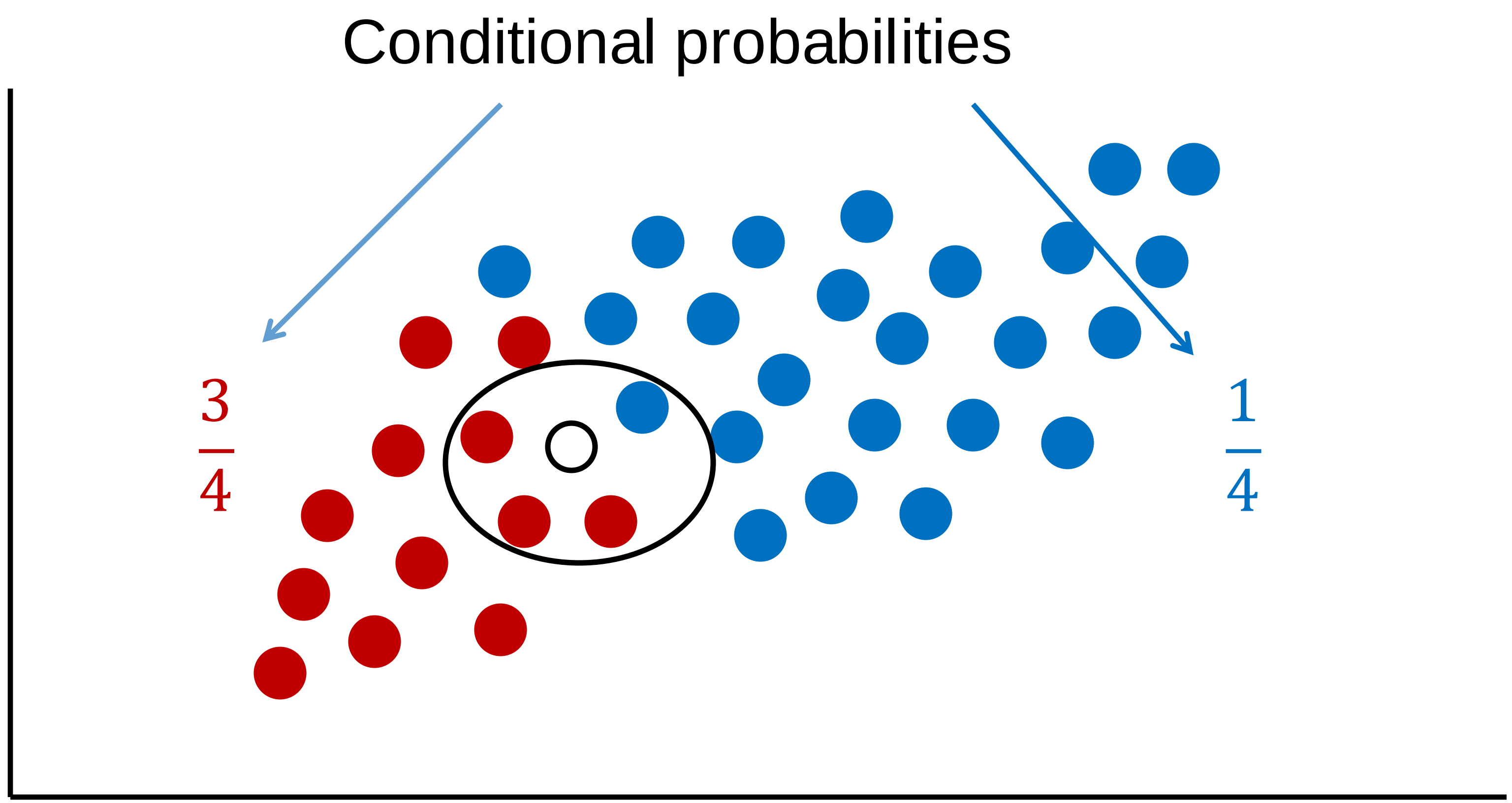


Naïve Bayes Classification

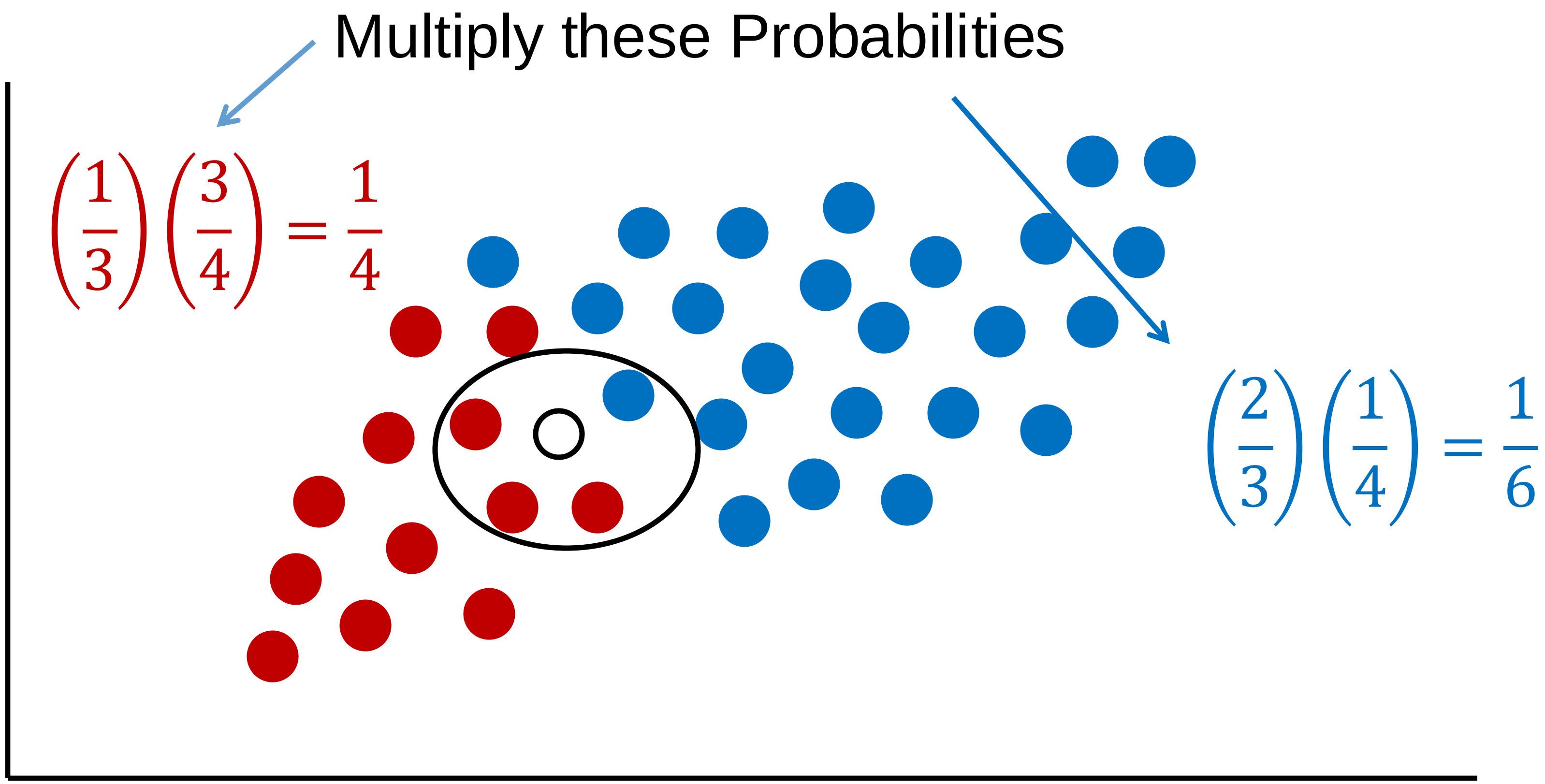
Take predefined closest number of observations.



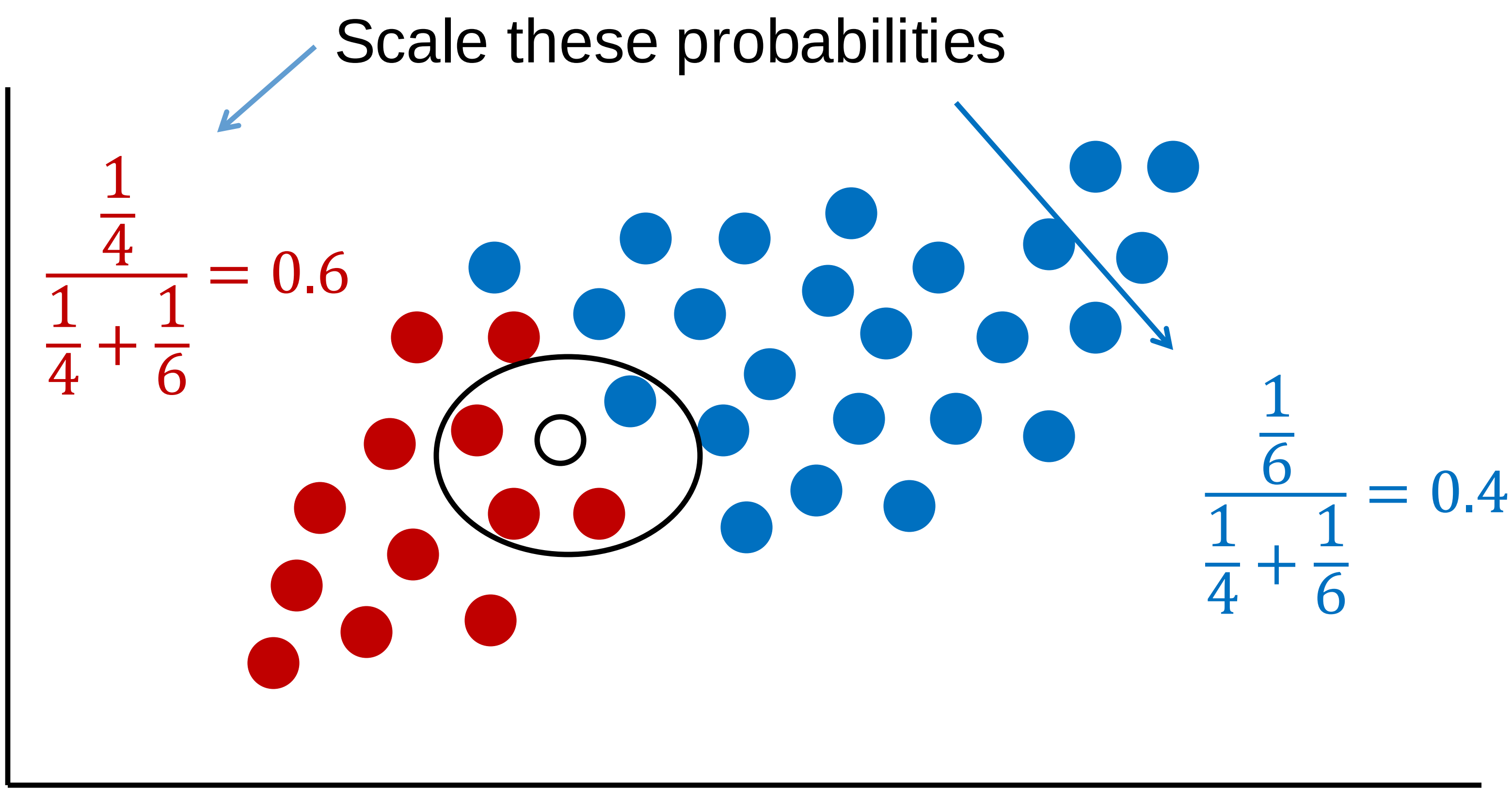
Naïve Bayes Classification



Naïve Bayes Classification



Naïve Bayes Classification



Naïve Bayes Assumption

- One of the big assumptions of the Naïve Bayes Classification method is one of the hardest things to accept:
  - Predictor variables are independent in their effects on the classification.
- This is a rather “naïve” assumption.
- Assumption doesn’t seem to bother posterior probabilities too greatly in case studies.

UNDERLYING MATH

Posterior Probabilities

- **Posterior probability →** Given values of variables for this observation, the predicted probability of success is…

$$P(success|variable\ values)$$

- **Prior probability →** Probability that an observation has those variable values.

Bayesian Classifiers

- Bayesian classifiers are based on Bayes' Theorem.
- **Naïve** Bayes Classifier assumes that the effect of the inputs are independent of one another.

- Bayes Theorem:

$$P(y_i|x_1, x_2, \dots, x_p) = \frac{P(y_i) \times P(x_1, x_2, \dots, x_p|y_i)}{P(x_1, x_2, \dots, x_p)}$$

Bayesian Classifiers

- Bayesian classifiers are based on Bayes' Theorem.
- **Naïve** Bayes Classifier assumes that the effect of the inputs are independent of one another.

- Remember the probabilities behind independent events:

$$P(A \cap B) = P(A) \times P(B)$$

$$P(A \cap B|C) = P(A|C) \times P(B|C)$$

Bayesian Classifiers

- Bayesian classifiers are based on Bayes' Theorem.
- **Naïve** Bayes Classifier assumes that the effect of the inputs are independent of one another.

- Bayes Theorem:

$$P(y_i \mid x_1, x_2, \ldots, x_p) = \frac{P(y_i) \times P(x_1 \mid y_i) \times \cdots \times P(x_k \mid y_i)}{P(x_1) \times \cdots \times P(x_k)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\,\&B) = \frac{P(Y) \times P(M\,|Y) \times P(B|Y)}{P(M) \times P(B)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\,\&B) = \frac{P(Y) \times P(M\,|Y) \times P(B|Y)}{P(M) \times P(B)}$$

Example

Size	Color	Accident
Large	**Blue**	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	**Blue**	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\,\&B) = \frac{P(Y) \times P(M\,|Y) \times \left(\frac{2}{6}\right)}{P(M) \times P(B)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\,\&B) = \frac{P(Y) \times \left(\frac{3}{6}\right) \times \left(\frac{2}{6}\right)}{P(M) \times P(B)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\ \&B) = \frac{P(Y) \times \left(\frac{3}{6}\right) \times \left(\frac{2}{6}\right)}{\left(\frac{3}{10}\right) \times P(B)}$$

Example

Size	Color	Accident
Large	**Blue**	Yes
Large	Red	Yes
Large	**Blue**	No
Large	**Blue**	No
Medium	Red	Yes
Medium	**Blue**	Yes
Medium	Red	Yes
Small	**Blue**	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\,\&B) = \frac{P(Y) \times \left(\frac{3}{6}\right) \times \left(\frac{2}{6}\right)}{\left(\frac{3}{10}\right) \times \left(\frac{5}{10}\right)}$$

Example

Size	Color	Accident
Large	Blue	**Yes**
Large	Red	**Yes**
Large	Blue	No
Large	Blue	No
Medium	Red	**Yes**
Medium	Blue	**Yes**
Medium	Red	**Yes**
Small	Blue	No
Small	Red	**Yes**
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y|M\ \&B) = \frac{\left(\frac{6}{10}\right) \times \left(\frac{3}{6}\right) \times \left(\frac{2}{6}\right)}{\left(\frac{3}{10}\right) \times \left(\frac{5}{10}\right)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of Yes, given both Medium and Blue.

$$P(Y \mid M \,\&B) = \frac{\left(\frac{6}{10}\right) \times \left(\frac{3}{6}\right) \times \left(\frac{2}{6}\right)}{\left(\frac{3}{10}\right) \times \left(\frac{5}{10}\right)} = \frac{2}{3}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P(N|M \,\&B) = \frac{P(N) \times P(M\,|N) \times P(B|N)}{P(M) \times P(B)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P(N|M\,\&B) = \frac{P(N) \times P(M\,|N) \times P(B|N)}{P(M) \times P(B)}$$

Example

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P\left(N \,\middle|\, M \,\&B\right) = \frac{P(N) \times \left(\frac{0}{4}\right) \times P(B|N)}{P(M) \times P(B)} = 0$$

☹

Problems of Zero Probability

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P\left(N \,\middle|\, M \,\&B\right) = \frac{P(N) \times \left(\frac{0}{4}\right) \times P(B|N)}{P(M) \times P(B)} = 0$$

- Predicted probability of zero regardless of everything else if certain values don’t occur for all levels of the outcome…
- Think “quasi-complete separation”ish …

Problems of Zero Probability

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P(N \mid M \,\&B) = \frac{P(N) \times \left(\frac{0}{4}\right) \times P(B|N)}{P(M) \times P(B)} = 0$$

	Yes	No
Small	1	2
Medium	3	0
Large	2	2

Laplace Correction (Laplace Estimator)

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P(N \mid M \,\&B) = \frac{P(N) \times \left(\frac{0}{4}\right) \times P(B|N)}{P(M) \times P(B)} = 0$$

	Yes	No
Small	1.01	2.01
Medium	3.01	0.01
Large	2.01	2.01

Laplace Correction (Laplace Estimator)

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P(N \mid M \,\&B) = \frac{P(N) \times \left(\frac{0.01}{4.03}\right) \times P(B|N)}{P(M) \times P(B)}$$

	Yes	No
Small	1.01	2.01
Medium	3.01	0.01
Large	2.01	2.01

Laplace Correction (Laplace Estimator)

Size	Color	Accident
Large	Blue	Yes
Large	Red	Yes
Large	Blue	No
Large	Blue	No
Medium	Red	Yes
Medium	Blue	Yes
Medium	Red	Yes
Small	Blue	No
Small	Red	Yes
Small	Red	No

- Compute the probability of No, given both Medium and Blue.

$$P(N \mid M \,\&B) = \frac{\left(\frac{4}{10}\right) \times \left(\frac{0.01}{4.03}\right) \times \left(\frac{3}{4}\right)}{\left(\frac{3}{10}\right) \times \left(\frac{5}{10}\right)}$$

$$= 0.005$$

Creating Output Probabilities

- The final probabilities will not likely sum to one so we force them to by dividing by their sum.

$$P\left(Y \,\middle|\, M\&B\right) = \frac{\left(\frac{2}{3}\right)}{0.005 + \left(\frac{2}{3}\right)} = 0.993$$

$$P\left(N \,\middle|\, M\&B\right) = \frac{0.005}{0.005 + \left(\frac{2}{3}\right)} = 0.007$$

NAÏVE BAYES IN R

Classification Target

Dr.LaBarr would only use Neive Bayes for a categorical output!!

- Inputs:
  - **Categorical variables** – determine probability based on cross-tabulation of each variable with target variable.
  - **Numerical variables** - determine probability based on either values from a Normal distribution with same mean and standard deviation as data OR kernel density estimation of the data.
- Output:
  - Probability that each observation belongs to each category of target variable.

Continuous Target

We should never use Neive Bayes for a Continuous Target!!!!

- Inputs:
  - **Categorical variables** – determine probability based on cross-tabulation of each variable with target variable.
  - **Numerical variables** - determine probability based on either values from a Normal distribution with same mean and standard deviation as data OR kernel density estimation of the data.
- Output:
  - **Value** of the target variable that is the highest probability.
  - Treats the continuous target as a large number of categories.

Ames Data

```
set.seed(4321)

training <- ames %>% sample_frac(0.7)
testing <- anti_join(ames, training, by = 'id')

training <- training %>%
  select(Sale_Price,
         Bedroom_AbvGr,
         Year_Built,
         Mo_Sold,
         Lot_Area,
         Street,
         Central_Air,
         First_Flr_SF,
         Second_Flr_SF,
         Full_Bath,
         Half_Bath,
         Fireplaces,
         Garage_Area,
         Gr_Liv_Area,
         TotRms_AbvGrd)
```

Naïve Bayes

```
set.seed(12345)
nb.ames <- naiveBayes(Sale_Price ~ ., data = training, laplace = 0, usekernel = TRUE)
```

Naïve Bayes – Tuning (ONLY CLASSIFICATION)

```
tune_grid <- expand.grid(
  usekernel = c(TRUE, FALSE),
  fL = c(0, 0.5, 1)
)

nb.ames.caret <- train(... ~ ., data = training,
                       method = "nb",
                       tuneGrid = tune_grid,
                       trControl = trainControl(method = 'cv', # Using 10-fold cross-validation
                                                number = 10))
```

SUMMARY

Naïve Bayes Summary

Advantages

- Simple to implement.
- Good at predictions.
  - **Especially good** classification for few categories.
- Perform best with categorical variables / text.
- Fast computational time.
- Robust to noise and irrelevant variables.

Disadvantages

- Independence assumption.
- Careful about normality assumption for continuous variables.
- Requires more memory storage than most models.
- Trust predicted categories more than probabilities.