Bayes' Rule

Foundations of Data Analysis

January 30, 2023

Naive Bayes Classifier

Iris species (versicolor, virginica, setosa)?

Naive Bayes Classifier

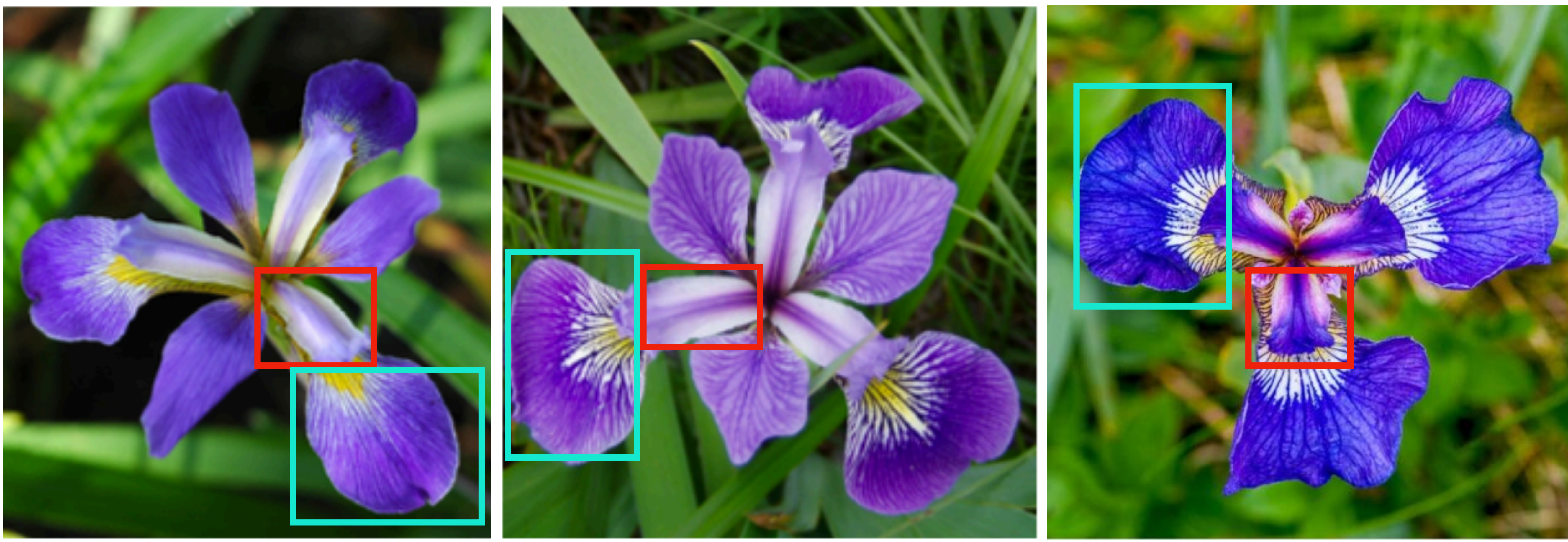
Iris virginica

Iris versicolor

Iris setosa

Naive Bayes Classifier

Iris virginica | Iris versicolor | Iris setosa



:Example regions of petal.

:Example regions of sepal.

Naive Bayes Classifier

- A probabilistic classifier:

  Features X and class C are random variables.

- Learn a probability distribution from the training data and predict results for testing (new) data.

$$P(C \mid X)$$

Naive Bayes Classifier

- A probabilistic classifier:

  Features X and class C are random variables.

- Learn a probability distribution from the training data and predict results for testing (new) data.

$$P(C\,|\,X)$$

For example: an iris test point X^* might give something like this:

C^*	setosa	versicolor	virginica
$P(C^* \mid X^*)$	0.80	0.15	0.05

Naive Bayes Classifier

- A **probabilistic** classifier:

  Features X and class C are random variables.

- Learn a probability distribution from the training data and predict results for testing (new) data.

$$P(C \mid X)$$

For example: an iris test point X^* might give something like this:

C^*	setosa	versicolor	virginica
$P(C^* \mid X^*)$	0.80	0.15	0.05

Naive Bayes Classifier

- What we want is:

$$P(C \mid X)$$

- What we can directly model from the data is:

$$P(X \mid C)$$

Bayes' Rule

Bayes' rule

Let's us "flip" a conditional:

$$p(B \mid A) = \frac{P(A \mid B) P(B)}{P(A)}$$

Deriving Bayes' Rule

Multiplication rule:

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

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

But these two equations are equal, so:

$$P(B \mid A)P(A) = P(A \mid B)P(B)$$

Dividing both sides by $P(A)$ gives us:

$$p(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Brain Teaser: Trick Coin

I have four coins. Three are normal, one side heads, one side tails. One is a trick coin where both sides are heads. I pick one coin at random and flip it. If it shows heads, what is the probability that it is the trick coin?

Trick Coin Example

$A =$ "heads", $B =$ "trick coin"

$$P(A \mid B) = 1.0$$

$$P(B) = 0.25$$

Trick Coin Example

A = "heads", B = "trick coin"

$$P(A \mid B) = 1.0$$

$$P(B) = 0.25$$

Consider two conditions: whether it is a trick coin or not

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$$
$$= 1.0 \times 0.25 + 0.5 \times 0.75 = \frac{5}{8}$$

$$p(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} = \frac{1.0 \times 0.25}{5/8} = \frac{2}{5} = 0.4$$

Random Variables

Definition

A **random variable** is a function defined on a sample space, Ω. Notation: $X : \Omega \to \mathbb{R}$

- A random variable is neither random or a variable.
- Just think of a random variable as assigning a number to every possible outcome.
- For example, in a coin flip, we might assign “tails” as 0 and “heads” as 1:

$$X(T) = 0, \quad X(H) = 1$$

Dice Example

Let $(\Omega, \mathscr{F}, P)$ be the probability space for rolling a pair of dice, and let X be the random variable that gives the sum of the numbers on the two dice. So,

$$X[(1,\ 2)] = 3, \quad X[(4,\ 4)] = 8, \quad X[(6,\ 5)] = 11$$

Even Simpler Example

Most of the time the random variable X will just be the identity function. For example, if the sample space is the real line, $\Omega = \mathbb{R}$, the identity function

$$
\begin{aligned}
&X : \mathbb{R} \to \mathbb{R}, \\
&X(s) = s
\end{aligned}
$$

is a random variable.

Defining Events via Random Variables

Setting a real-valued random variable to a value or range of values defines an event.

$$[X = x] = \{s \in \Omega : X(s) = x\}$$

$$[X < x] = \{s \in \Omega : X(s) < x\}$$

$$[a < X < b] = \{s \in \Omega : a < X(s) < b\}$$

Joint Probabilities

Two binary random variables:

C = cold / no cold = (1/0)

R = runny nose / no runny nose = (1/0)

Event $[C = 1]$: “I have a cold”

Event $[R = 1]$: “I have a runny nose”

Joint event

$[C = 1] \cap [R = 1]$: “I have a cold and a runny nose”

Notation for joint probabilities:

$$P(C = 1, R = 1) = P([C = 1] \cap [R = 1])$$

Cold Example: Probability Tables

Two binary random variables:

C = cold / no cold = (1/0)

R = runny nose / no runny nose = (1/0)

Joint probabilities:

		C = 0	C = 1
R	0	0.50	0.05
	1	0.20	0.25

Cold Example: Marginals

		C = 0	C = 1
R	0	0.50	0.05
	1	0.20	0.25

Marginals:

$$P(R = 0) = 0.55,\ P(R = 1) = 0.45$$

$$P(C = 0) = 0.70,\ P(C = 1) = 0.30$$

Cold Example: Conditional Probabilities

		C = 0	C = 1	
R	0	0.50	0.05	0.55
	1	0.20	0.25	0.45
		0.7	0.3	

Conditional Probabilities:

$$P(C = 0 \mid R = 0)$$

$$P(C = 1 \mid R = 1)$$

Cold Example: Conditional Probabilities

		C = 0	C = 1	
R	0	0.50	0.05	0.55
	1	0.20	0.25	0.45
		0.7	0.3	

Conditional Probabilities:

$$P(C=0 \mid R=0) = \frac{P(C=0, R=0)}{P(R=0)} = \frac{0.50}{0.55} \approx 0.91$$

$$P(C=1 \mid R=1) = \frac{P(C=1, R=1)}{P(R=1)} = \frac{0.25}{0.45} \approx 0.56$$

Cold Example

		C = 0	C = 1	
R	0	0.50	0.05	0.55
	1	0.20	0.25	0.45
		0.7	0.3	

Remember:

$$P(C) = 0.3$$

$$P(C \mid R) = 0.56$$

What if I didn't give you the full table, but just:

$$P(R \mid C) = 0.83 \; > \; P(R) = 0.45$$

What can you say about the increase

$P(C \mid R) > P(C)$?

Cold Example

Notice, having a cold *increases* my chance for a runny nose by the factor,

$$\frac{P(R \mid C)}{P(R)} = \frac{0.83}{0.45} = 1.85$$

How does such a ratio increase if I flip the conditional?

$$\frac{P(C \mid R)}{P(C)} = \frac{P(C \cap R)}{P(R)P(C)} = \frac{P(R \mid C)}{P(R)} = 1.85$$