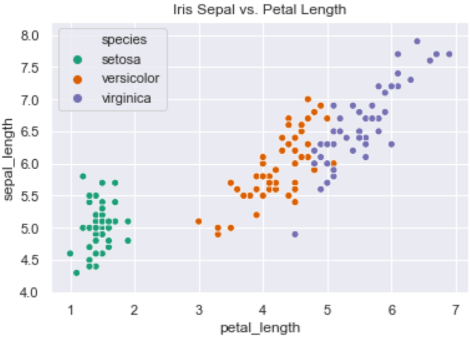
Classification and Naïve Bayes

Foundations of Data Analysis

January 31, 2019

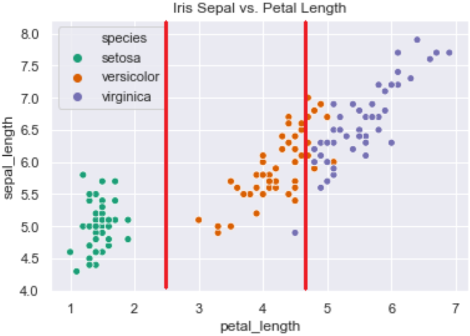
Classification

Say we want to automatically identify an iris species based on its petal and sepal length measurements.



A Classifier is a *Decision Rule*

`x` = “petal length”, `c` = “species”




```
if x < 2.5 :  c = 'setosa'
if 2.5 < x < 4.7 :  c = 'versicolor'
if x > 4.7 :  c = 'virginica'
```

Classification Task

Training:
Learn a decision rule, based on training data, to predict a class C from features X.

Testing:
Use trained classifier to predict unknown class C^* from features of new testing data, X^*.

Classification Task

Training:
Learn a decision rule, based on training data, to predict a class C from features X.

Testing:
Use trained classifier to predict unknown class C^* from features of new testing data, X^*.

Important! Training and testing data should be completely separate!

Probabilistic Classifier

Features X and class C are random variables.

Learn a probability distribution from the training data:

$$P(C \mid X)$$

Probabilistic Classifier

Features X and class C are random variables.

Learn a probability distribution from the training data:

$$P(C \mid X)$$

Imaginary 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

Bayes' Rule for Classification

$$P(C \mid X) = \frac{P(X \mid C)P(C)}{P(X)}$$

$P(X \mid C)$ **Likelihood** - learned from data

$P(C)$ **Prior** - determined beforehand

$P(X)$ **Evidence** - not needed for decision

Naïve Bayes

Multidimensional features $X = (X_1, X_2, \ldots, X_d)$

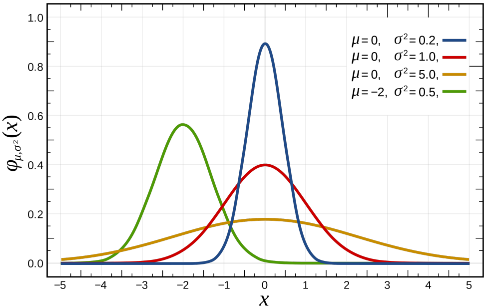
Naïve Bayes

Multidimensional features $X = (X_1, X_2, \ldots, X_d)$

"Naïve" Assumption:
Assume features X_i are independent, given the class C:

$$P(X \mid C) = P(X_1 \mid C) \times P(X_2 \mid C) \times \cdots \times P(X_d \mid C)$$

Gaussian or Normal Distribution



Probability density function:

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Notation: $x \sim N(\mu, \sigma^2)$

Mean, μ, and variance, σ^2, are parameters.

See `https://en.wikipedia.org/wiki/Normal_distribution`

How to “Train” a Normal Distribution

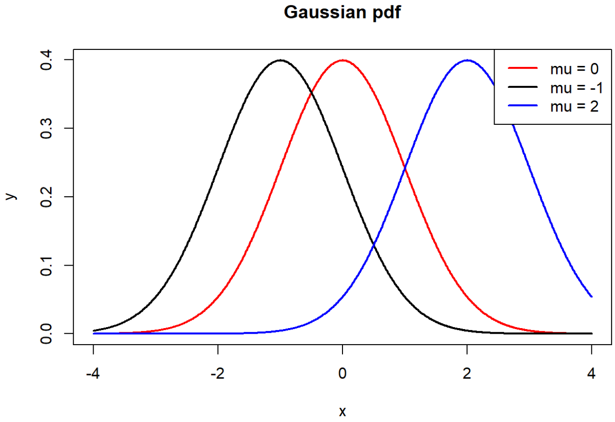
Given training data: $x_1, x_2, \ldots, x_n$

How to “Train” a Normal Distribution

Given training data: $x_1, x_2, \ldots, x_n$

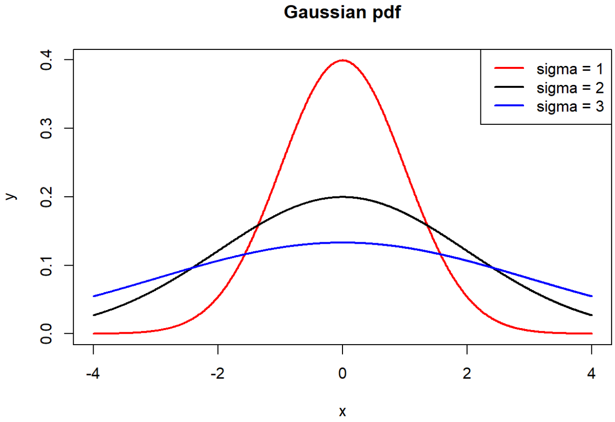
Set parameters:
Mean: $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$
Variance: $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$

Gaussian μ Parameter



Shifts the pdf, shape stays the same

Gaussian σ Parameter



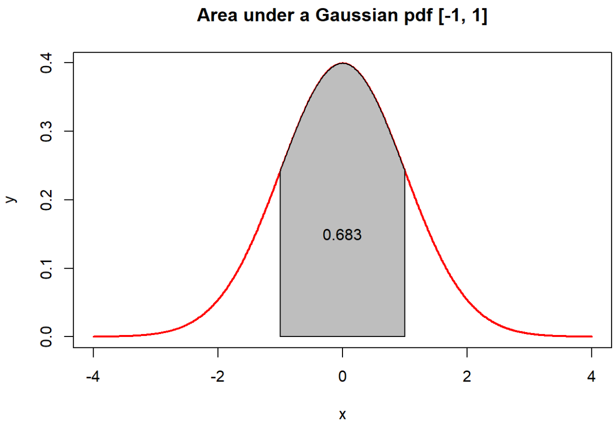
Stretches/shrinks the pdf, position stays the same

Probabilities of Continuous Random Variables

Probability is given by area under the pdf:

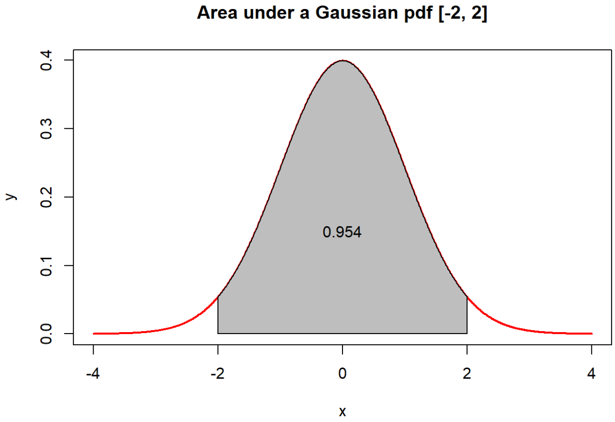
$$P(a < X < b) = \int_a^b p(x)dx$$

Gaussian Area



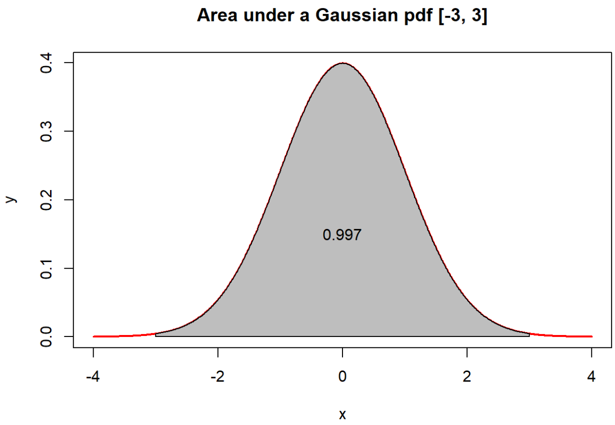
Units of horizontal axis are σ

Gaussian Area



Units of horizontal axis are σ

Gaussian Area



Units of horizontal axis are σ