

Classification and Naïve Bayes

Foundations of Data Analysis

February 8, 2022

Brain Teaser

You have two black socks, two white socks, two red socks, and two blue socks in your sock drawer. If you pick two socks out at random (without looking!), what is the probability that they match?

Irises



iris virginica



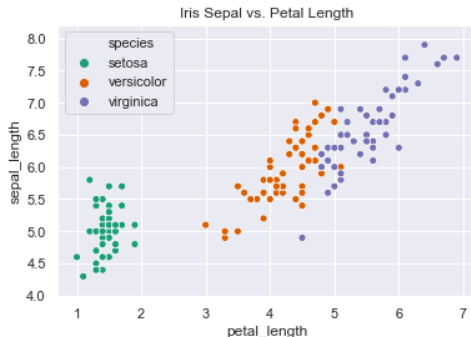
iris versicolor



iris setosa

Classification

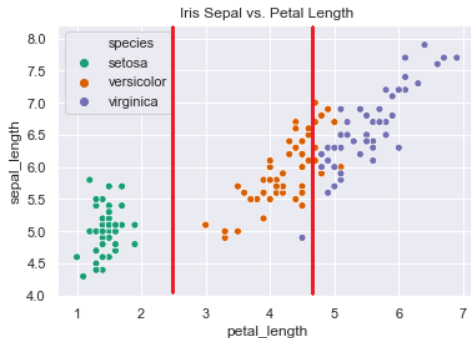
Say we want to automatically identify an iris species based on its petal and sepal length measurements.



This is a famous data set in machine learning / statistics, from Ronald Fisher in 1936!

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^* .

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:

$$P(C | X)$$

Imaginary Example:

An iris test point X^* might give something like this:

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

Bayes' Rule for Classification

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

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

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

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

Naïve Bayes

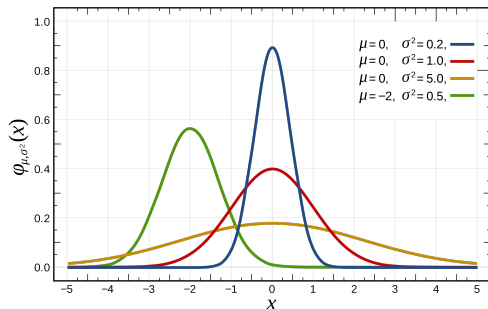
Multidimensional features $X = (X_1, X_2, \dots, X_d)$

“Naïve” Assumption:

Assume features X_i are independent, given the class C :

$$P(X | C) = P(X_1 | C) \times P(X_2 | C) \times \dots \times P(X_d | C)$$

Gaussian or Normal Distribution



Probability density function (pdf):

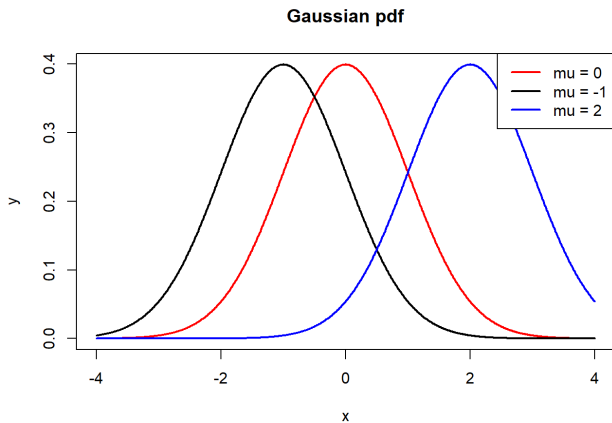
$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

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

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

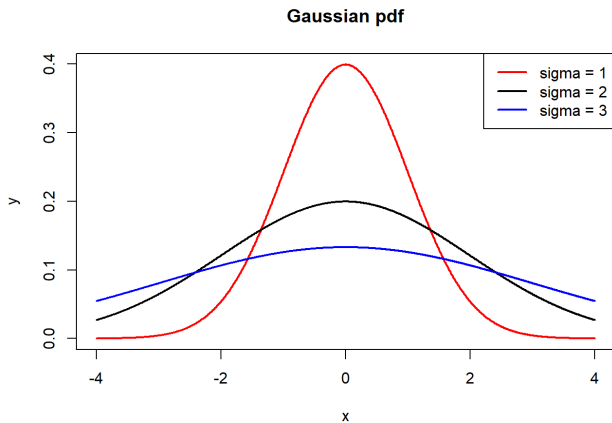
See https://en.wikipedia.org/wiki/Normal_distribution

Gaussian μ Parameter



Shifts the pdf, shape stays the same

Gaussian σ Parameter



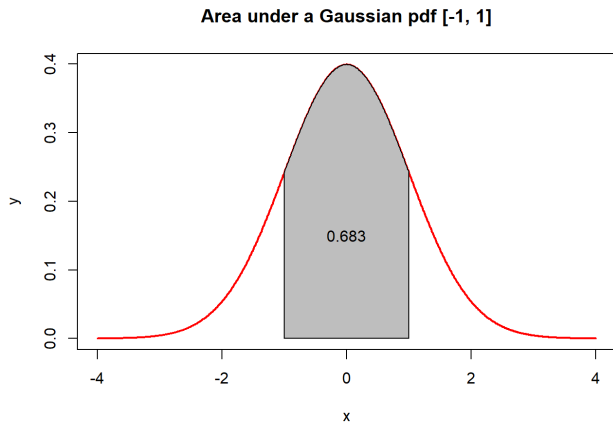
Stretches/shrinks the pdf, position stays the same

Probabilities of Continuous Random Variables

Probability is given by area under the pdf:

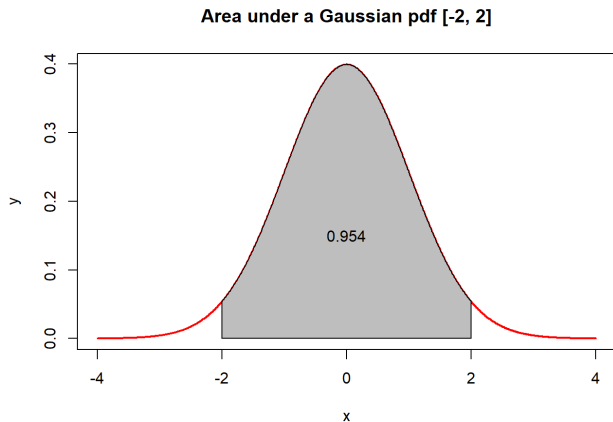
$$P(a < X < b) = \int_a^b p(x)dx$$

Gaussian Area



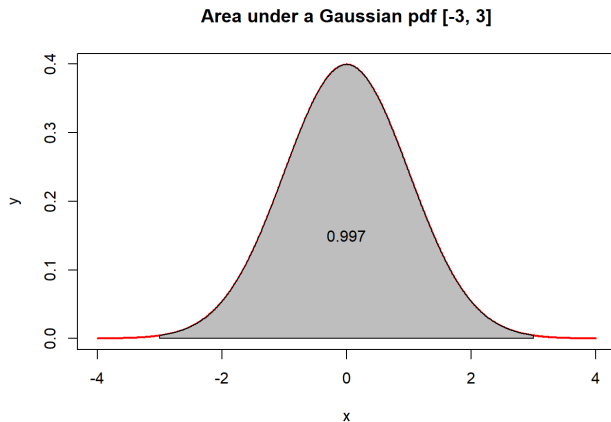
Units of horizontal axis are σ

Gaussian Area



Units of horizontal axis are σ

Gaussian Area



Units of horizontal axis are σ

Gaussian Naïve Bayes

Likelihood is Gaussian pdf:

$$p(x | C = c_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x - \mu_k)^2}{2\sigma_k^2}\right)$$

- ▶ The Gaussian depends on the class
 $C \in \{c_1, c_2, \dots, c_K\}$
- ▶ Each class needs a mean, μ_k , and a variance, σ_k^2

How to “Train” a Gaussian Distribution

For each feature in your data, given training data:
 x_1, x_2, \dots, x_n all from the k th class

Set parameters:

Mean: $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n x_i$

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

Evidence Calculation

If we have K classes $C \in \{c_1, c_2, \dots, c_K\}$:

$$p(x) = \sum_{k=1}^K p(X | C = c_k)P(C = c_k),$$

using Total Probability.

For the case that we have two classes:

$$p(x) = p(x | C = c_1)P(C = c_1) + p(x | C = c_2)P(C = c_2)$$

Choosing a Prior

How to set $P(C = c_k)$?

- ▶ Equally likely: $P(C = c_k) = \frac{1}{K}$
- ▶ Frequency of classes in training data
- ▶ Derive from previous experiments or knowledge

Putting It All Together

- ▶ Pick a prior: $P(C = c_k)$
- ▶ Train your Gaussians on training data: μ_k, σ_k^2
- ▶ For each test data point, $x^* = (x_1^*, x_2^*, \dots, x_d^*)$, compute the likelihood:

$$p(x^* | C = c_k) = p(x_1^* | C = c_k) \times p(x_2^* | C = c_k) \times \dots \times p(x_d^* | C = c_k)$$

- ▶ Compute the class probabilities:

$$P(C = c_k | x^*) = \frac{p(x^* | C = c_k)P(C = c_k)}{p(x^*)}$$

- ▶ Classify x^* as the class c_k with highest probability