Bayesian learning

We formulate classification as a statistical problem and questions such as "What is the probability of this sample begin from class A?".

REV. T. BAYES

Bayes' theorem

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- $P(h)$ = prior probability of hypothesis h (e.g. a class)
- $P(D)$ = prior probability of observation (data) D
- $P(h|D)$ = probability of c given D
- $P(D|h)$ = probability of D given h

  D is your measurement or data!

Classification

We want to select the class with the highest Bayes (a posteriori) probability given the measurement D.

Maximum a posteriori hypothesis h_{MAP}:

$$\begin{aligned} h_{MAP} &= \arg\max_{c \in C} P(c|D) \\ &= \arg\max_{c \in C} \frac{P(D|c)P(c)}{P(D)} \\ &= \arg\max_{c \in C} P(D|c)P(c) \end{aligned}$$

Maximum likelihood vs. maximum a posteriori

If assume equal priories, $P(h_i) = P(h_j)$, then the optimal decision simplies to the *maximum likelihood* (ML) classification

$$h_{ML} = \arg\max_{h_i \in C} P(D|h_i)$$

Example

A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, $.008$ of the entire population have this cancer.

$$\begin{aligned} P(cancer) &= & P(\neg cancer) &= \\ P(+|cancer) &= & P(-|cancer) &= \\ P(+|\neg cancer) &= & P(-|\neg cancer) &= \end{aligned}$$

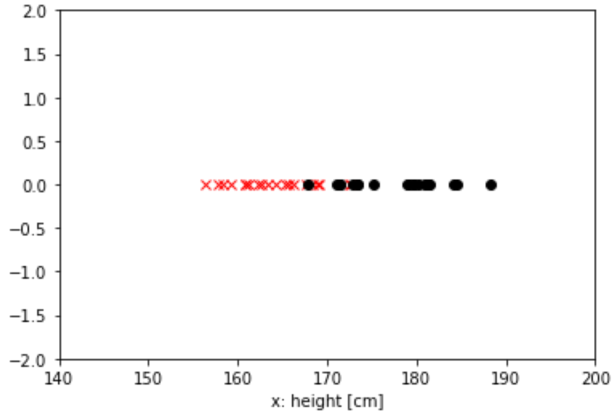
h_{MAP}? h_{ML}?

Basic formulas for probabilities

- Product Rule: probability $P(A \wedge B)$ of a conjunction of two events A and B:
$$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$
- Sum Rule: probability of a disjunction of two events A and B:
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$
- Theorem of total probability: if events $A_1, \ldots, A_n$ are mutually exclusive with $\sum_{i=1}^{n} P(A_i) = 1$, then
$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

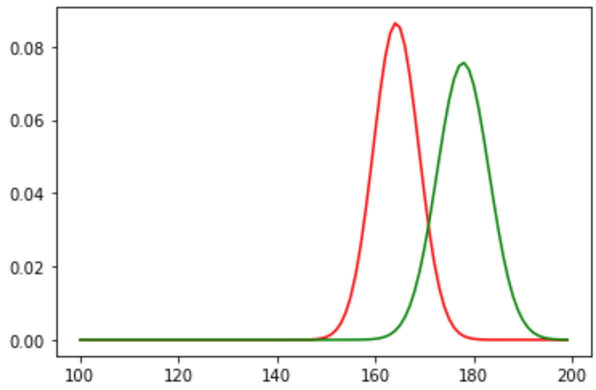
Example 2 Let's study Bayes probabilities for the same data we used for classification in the previous lectures: male/female height measurements.

In [1]:
```
# Let's have a bayesien look on the male/female classification problem
import matplotlib.pyplot as plt
import numpy as np
plt.xlabel('x: height [cm]')
plt.axis([140,200,-2,2])
x_1 = np.random.normal(165,5,20) # Measurements from the class 1
x_2 = np.random.normal(180,6,20) # Measurements from the class 2
plt.plot(x_1,np.zeros(len(x_1)),'rx')
plt.plot(x_2,np.zeros(len(x_2)),'ko')
plt.show()
```



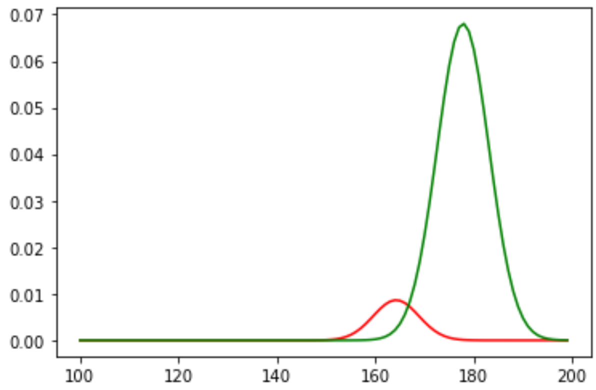
In [2]:
```
# Let's model the measurement data as Gaussians (normal distribution)
import scipy.stats as stats
mu1 = np.mean(x_1)
mu2 = np.mean(x_2)
sigma1 = np.std(x_1)
sigma2 = np.std(x_2)
x = np.arange(100,200,1)
plt.plot(x, stats.norm.pdf(x, mu1, sigma1),'r-')
plt.plot(x, stats.norm.pdf(x, mu2, sigma2),'g-')
print([mu1,mu2,sigma1,sigma2])
```

```
[164.21296544094864, 177.8244186313693, 4.614654891110036, 5.276973047307949]
```



In [3]:
```
# The same plot with priors (change priori_1 to understand its meaning)
priori_1 = 0.1
priori_2 = 1-priori_1
plt.plot(x, priori_1*stats.norm.pdf(x, mu1, sigma1),'r-')
plt.plot(x, priori_2*stats.norm.pdf(x, mu2, sigma2),'g-')
print([mu1,mu2,sigma1,sigma2])
```

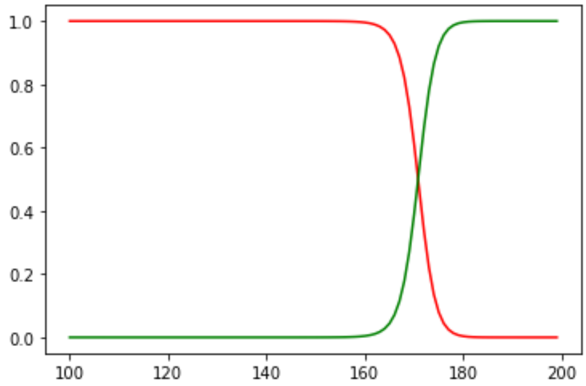
```
[164.21296544094864, 177.8244186313693, 4.614654891110036, 5.276973047307949]
```



In [4]:
```
# Let's study posteriori probabilities using Bayes rule
priori_1 = 0.5
priori_2 = 1.0-priori_1
# I use the Bayes theorem to calculate the posteriori probabilities
posteriori_1 = stats.norm.pdf(x, mu1, sigma1)*priori_1/(stats.norm.pdf(x, mu1,
sigma1)*priori_1+stats.norm.pdf(x, mu2, sigma2)*priori_2)
posteriori_2 = stats.norm.pdf(x, mu2, sigma2)*priori_2/(stats.norm.pdf(x, mu1,
sigma1)*priori_1+stats.norm.pdf(x, mu2, sigma2)*priori_2)
x_d = np.argmin(np.abs(posteriori_1-posteriori_2)) # note: only approximate po
int
print(f"Decision boundary at {x[x_d]} cm")
plt.plot(x, posteriori_1,'r-')
plt.plot(x, posteriori_2,'g-')
```

```
Decision boundary at 171 cm
```

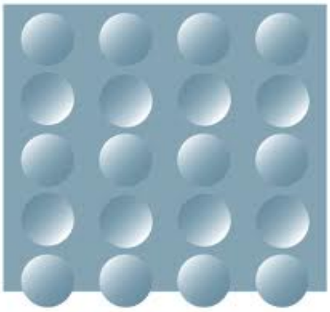
Out[4]:
```
[<matplotlib.lines.Line2D at 0x7f1fc2044220>]
```

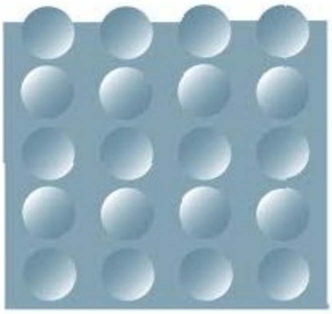


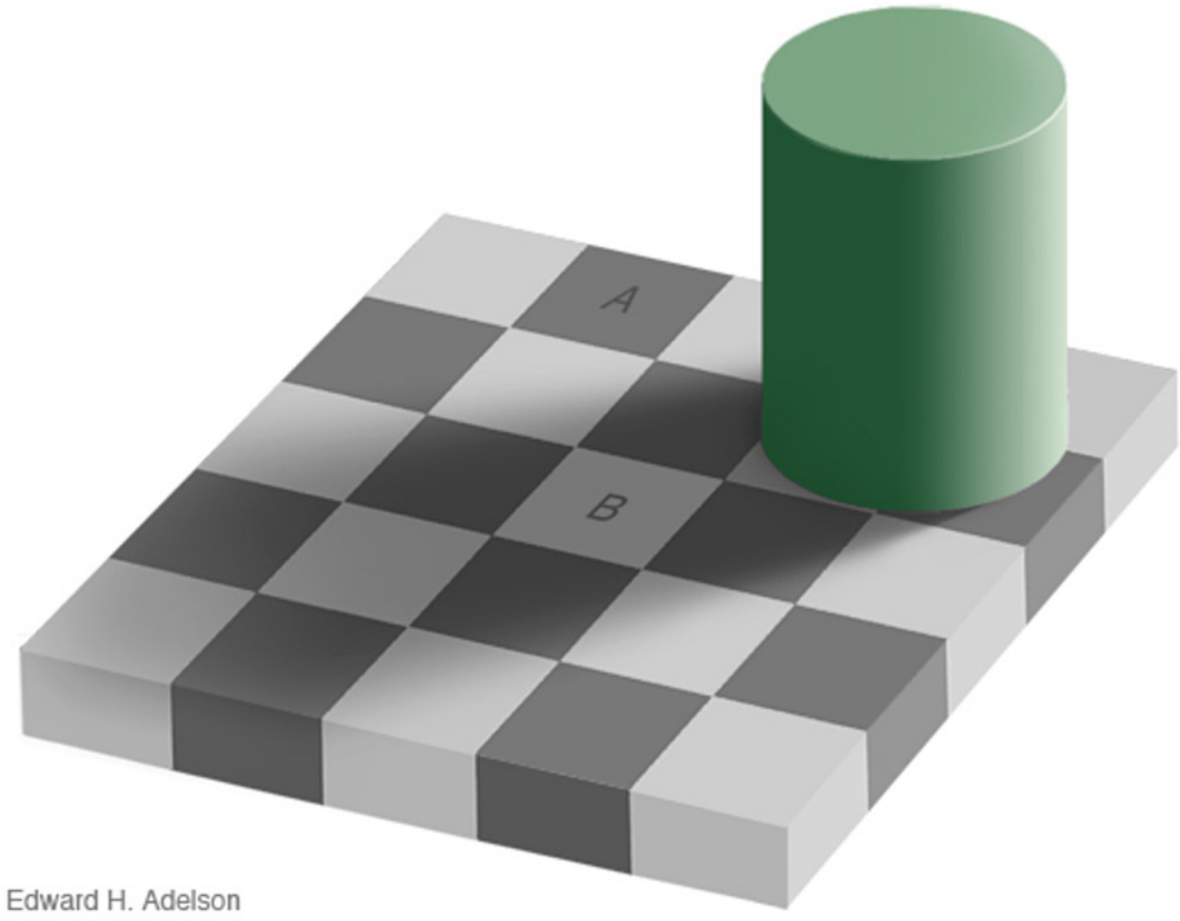
Bayesian thinking

Since Bayes rule guarantees optimal descision making for the given hypothesis space, are priors part of human brains as well?

Bayes and computer (human) vision

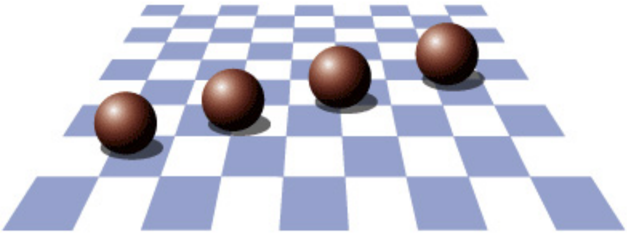


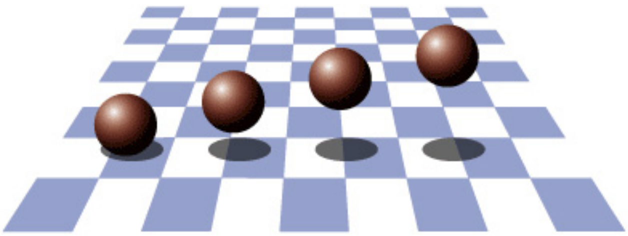



A
B
Edward H. Adelson

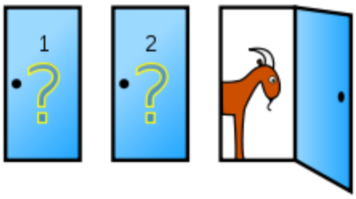

A


B





Probabilities are not necessarily intuitive - difficult magic to handle



The Monty Hall problem

Bayes regression

To understand how Bayes theorem can be used in regression it is better to first talk about estimation. The estimate of probability for a binary variable (coin tossing) is familiar to all of us

$$\mu = \frac{n_c}{n}$$

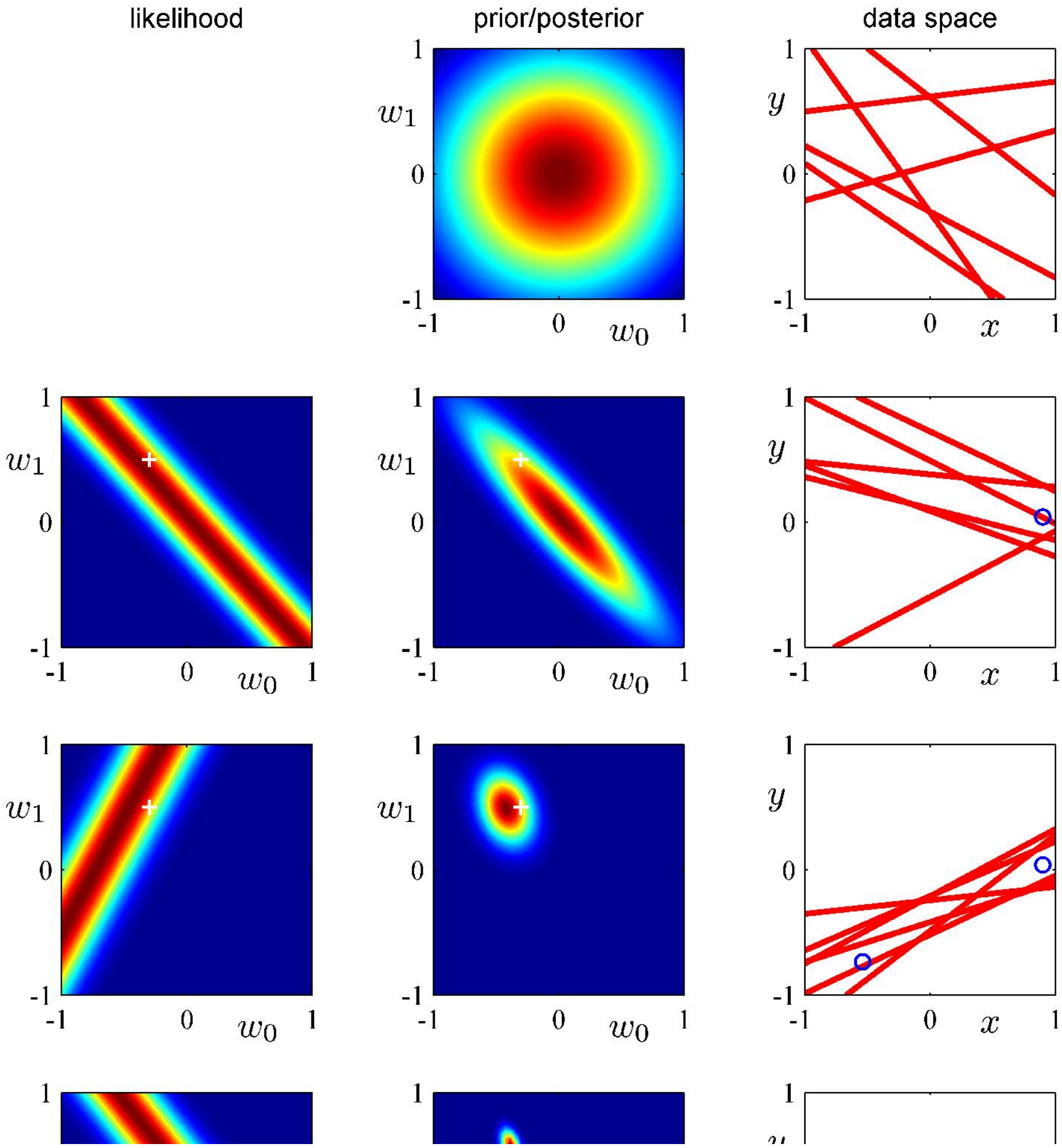
where μ is the probability of the event 1 (e.g. heads) and naturally $1-\mu$ is the probability of the event 2 (tails). n_c is the number of positive outcomes (heads) and n is the total number of samples. This solution btw. is μ_{ML}.

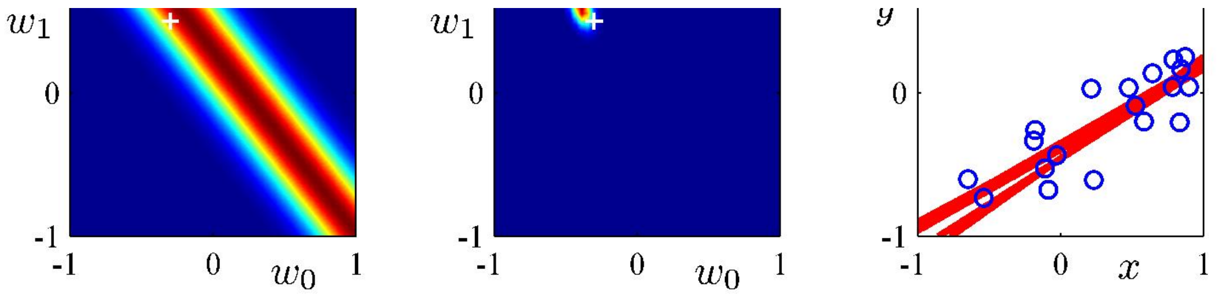
The bayesian (Maximum A Posteriori) estimate of μ_{MAP} is

$$\mu = \frac{n_c + mp}{n+m}$$

where is m additional "virtual tosess" and p is the prior estimate of μ i.e. how many of the virtual tosses land as heads. It is important to notice that $\mu_{MAP} = \mu_{ML}$ when $n \to \infty$.

Now you need to think how Maximum a posteriori approach could be realized for linear regression?





Bayesian thinking needs reshaping many statistical principles and is a powerful tool (Gelman et al. 2013). See also Aki Vehtari Web site (Aalto University) for links to software tools and other materials ([https://users.aalto.fi/~ave/ (https://users.aalto.fi/~ave/)]).

References

C.M. Bishop (2006): Pattern Recognition and Machine Learning, Chapter 1-2.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari and D.B. Rubin (2013): Bayesian Data Analysis, Third Edition, Chapman and Hall/CRC. [https://users.aalto.fi/~ave/BDA3.pdf (https://users.aalto.fi/~ave/BDA3.pdf)]