Types of events:
Prior - probability of an event to happen.
E.g. probability of selecting a fish S is P(S) = 0.4, selecting a fish B is P(B) = 0.6

Evidence - probability of measuring a certain value of features from all the samples.
E.g. probability of finding the evidence width of a fish = 26cm => P(width=26cm) = 0.3

Class-conditional - probability of measuring a certain feature value from samples that belong to a certain class.
E.g. a fish from the class salmon - probability of measuring a width of 26cm, given that we have a salmon => P(width=26cm | salmon) = P(width=26cm AND salmon) / P(S) = 0.5

Posterior - probability of a sample to belong to a certain class given that we have measured a certain feature value

Posterior probability with the ***Bayes formula***

Joint probability
Probability of two events occurring at the same time (e.g. measuring features **x** and class **C** of a sample)

$$P(x, C) = P(C|x)P(x)$$
$$P(x, C) = P(x|C)P(C)$$

$$P(C|x)P(x) = P(x|C)P(C)$$

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

posterior: $P(C|x)$; class-conditional: $P(x|C)$; prior: $P(C)$; evidence: $P(x)$

The evidence is computed as: $P(X) = P(X|C_1)P(C_1) + P(X|C_2)P(C_2) + \ldots$

Evidence = P(X,a) = P(X,a|True)P(True) + P(X,a|False)P(False) +..

Naive Bayes - Estimating probability
P(True|X,a) = (P(X,a|True)P(True)) / P(X,a)
=>(P(X|True)P(a|True)P(True)) / P(X,a)

Bernoulli Trial: formula

From N choose k $\binom{N}{k} = \frac{N!}{k!(N-k)!}$ *Binomial coefficient*

Probability of ***x=k*** successes given **N** trials (where ***p*** is the probability of a single success)

$$P(x = k)|_N = \binom{N}{k} p^k (1-p)^{N-k}$$

Binomial distribution

Univariate: we consider only one dimension (on feature)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

The Gaussian function has two parameters that can be estimated:

mean: $\mu = \frac{1}{n}\sum_{k=1}^{n} x_k$

variance: $\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu)^2$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

Probabilistic dichotomizer

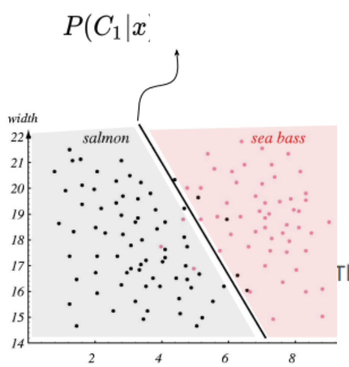
We use the Posterior probability as discriminant function.

We define the classifier as:
$g(x) = g_1(x) - g_2(x) = P(C_1|x) - P(C_2|x)$

And the decision regions become:

$g(x) > 0 \iff P(C_1|x) > P(C_2|x)$ — **We decide for class 1**

$g(x) < 0 \iff P(C_1|x) < P(C_2|x)$ — **We decide for class 2**



Example: the posterior probability

Estimation of the posterior probability using the Bayes rule (and the quantities we have computed earlier):

$$P(C_1|x) = \frac{P(C_1)P(x|C_1)}{P(x)} = \frac{0.4\frac{1}{10\sqrt{2\pi}}e^{-\frac{x^2}{2\cdot 10^2}}}{P(x)}$$

$$P(C_2|x) = \frac{P(C_2)P(x|C_2)}{P(x)} = \frac{0.6\frac{1}{20\sqrt{2\pi}}e^{-\frac{(x-35)^2}{2\cdot 20^2}}}{P(x)}$$

Note: for a new data point **x'** we can compute the posterior probability of it belonging to class 1 and the posterior probability of belonging to class 2. We select the class with highest posterior probability.

Detection - Recognise the presence of certain samples
Classification - Recognise a sample and assign it to a class
Identification - Recognise a pattern and give a precise unique identity to it
Authorisation - After identification, you assign privileges or rights to the person (recognised pattern)

Example: the class-conditional probability

aussian functions to model the class-conditional probability.
have mean and standard deviation of the two classes!!!

$\mu_1 = 0, \ \sigma_1^2 = 10^2$
$\mu_2 = 35, \sigma_2^2 = 20^2$

$$P(x|C_1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = \frac{1}{10\sqrt{2\pi}} e^{-\frac{x^2}{2\cdot 10^2}}$$

$$P(x|C_2) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} = \frac{1}{20\sqrt{2\pi}} e^{-\frac{x^2-35}{2\cdot 20^2}}$$

Computing the decision boundary (2)

We can use the *natural logarithm* function ln(x) to further simplify the equation:

$$ln\left[8e^{-\frac{x^2}{2\cdot 10^2}}\right] = ln\left[6e^{-\frac{(x-35)^2}{2\cdot 20^2}}\right]$$

$$ln\,[8] + ln\left[e^{-\frac{x^2}{2\cdot 10^2}}\right] = ln\,[6] + ln\left[e^{-\frac{(x-35)^2}{2\cdot 20^2}}\right]$$

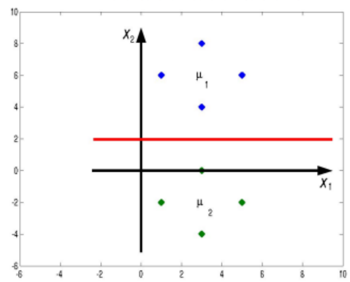
$$ln8 - \frac{x^2}{2\cdot 10^2} = ln6 - \frac{(x-35)^2}{2\cdot 20^2}$$

And reduce it to: $3x^2 + 70x - 1445.14 = 0$

Let us see it with an example

$\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \sigma_1 = \sigma_2 = \sqrt{2} \quad P(C_1) = P(C_2) = 0.5$

Class 1: {(3,8), (1,6), (5,6), (3,4)}
Class 2: {(3,0), (1,-2), (3,-4), (5,-2) }



To determine the decision boundary we have to compute:

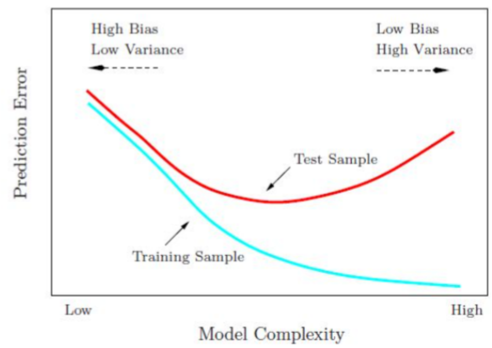
$$g_1(\mathrm{x}) = g_2(\mathrm{x})$$

$\Rightarrow \quad -\frac{\|x-\mu_1\|^2}{2\sigma^2} + lnP(C_1) = -\frac{\|x-\mu_2\|^2}{2\sigma^2} + lnP(C_2)$

$\Rightarrow \quad (\mathrm{x} - \mu_1)^t(\mathrm{x} - \mu_1) = (\mathrm{x} - \mu_2)^t(\mathrm{x} - \mu_2)$

$\Rightarrow \quad [x_1 - 3 \quad x_2 - 6]\begin{bmatrix} x_1 - 3 \\ x_2 - 6 \end{bmatrix} = [x_1 - 3 \quad x_2 + 2]\begin{bmatrix} x_1 - 3 \\ x_2 + 2 \end{bmatrix}$

$\Rightarrow \quad (x_1-3)^2 + (x_2-6)^2 = (x_1-3)^2 + (x_2+2)^2 \quad \Rightarrow \quad \boxed{x_2 = 2}$



Accuracy: rate of correct classifications

$$Acc = \frac{TP+TN}{TP+FP+TN+FN}$$

		Predicted class	
		Positive	**Negative**
True class (ground truth)	**Positive**	TP	FN
	Negative	FP	TN

Precision: rate of correct positive out of all classified positive

$$Pr = \frac{TP}{TP+FP}$$

Recall: rate of correct positive out of all ground truth positives

$$Re = \frac{TP}{TP+FN}$$