### What is Machine Learning?

$$ \text{data} + \text{model} \rArr \text{prediction}$$

* : observations, could be actively or passively acquired (meta-data).
* : assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
* : an action to be taken or a categorization or a quality score.

### Two important Gaussian Properties

* Sum of Gaussianv
* Sum of Gaussian variables is also Gaussian.

And the sum is distributed as

Aside: As sum increase, sum of non-Gaussian, finite variance variables is also Gaussian because of **central limit theorem**. - Scaling a Gaussian Scaling a Gaussian leads to a Gaussian.

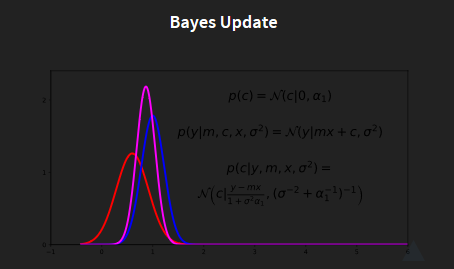
The **central limit theorem** (CLT) establishes that, in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (informally a “bell curve”) even if the original variables themselves are not normally distributed.

### Prior Distribution

* Bayesian inference requires a prior on the parameters.
* The prior represents your belief *before* you see the data of the likely value of the parameters.
* For linear regression, consider a Gaussian prior on the intercept:

### Posterior Distribution

* Posterior distribution is found by combining the prior with the likelihood.
* Posterior distribution is your belief *after* you see the data of the likely value of the parameters.
* The posterior is found through **Bayes’ Rule**
* The likelihood is not a density over , it’s a function of . Here, is a parameter of this density. The normalization step, e.g. find the suitable way to compute the is the most difficult setp.



Bayes Update

The red line describe the probablity for with:

the blue line stands for the observation, with:

Note that, this is a likelihood function over not a distribution. based on the bayes’rule, the posterior could be written as:

Math Trick:

complete the square of the quadratic form to obtain

where

and

.

Piror comes from the model, where we think about it. And likelihood is coming from the data.

### Stages to Derivation of the Posterior

* Multiply likelihood by prior
  + they are “exponentiated quadratics”, the answer is always also an exponentiated quadratic because
* Complete the square to get the resulting density in the form of a Gaussian.
* Recognise the mean and (co)variance of the Gaussian. This is the estimate of the posterior.

### Multivariate Regression Likelihood

* Noise corrupted data point
* Multivariate regression likelihood:
* Multivariate Gaussian piroi:

The independent multivariate Gaussian could be seen as the independent Gaussian and multiple them and rotate the results.

#### Independent Gaussians:

and we can write it with linear algebra form:

and then, rename it:

means the determinant of the matrix.

#### Correlated Gaussian

Form correlated from original by rotating the data space using matrx R.

this gives a covariance matric:

which in some view is the result of the principal compoment.

### Multivariate Consequence

if

$$x \sim \mathcal{N}(\mu,\varSigma)$$

and

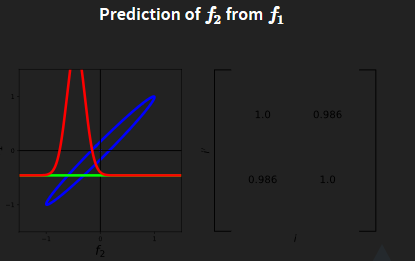
then

$$y\sim \mathcal{N}(W\mu,W\varSigma W^T)$$

we can say the first equaption is the prior of , the second is the likelihood, the last is the marginal of . If we set $\mu = 0, \varSigma = I$, so it is just the inverse of PCA.

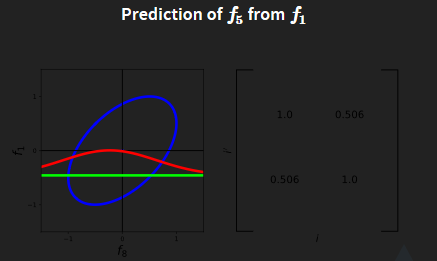
## Prediction with Correlated Gaussians

* Prediction of from requires multivariate *conditional density*.
* Multivariate conditional density is *also* Gaussian.
* Here covariance of joint density is given by

  
Take the picture as the example:

since that example is in 1D, all the values are scalar. , , so

the variance is 0.0396, and the standard variance is 0.2.



Prediction