Some random formulas and procedures which may be of use. There’s a lot to add (feel free to contribute)

## Probability

* (same for a multivariate case)
* . for constant a
* In a multivariate context, , where is the covariance matrix, also defined as .
* . In a multivariate context,
* Conditional probability: .
* Bayes’ theorem (where is the number of options and we want the th case)
* Gaussian formulas:

A picture containing diagram

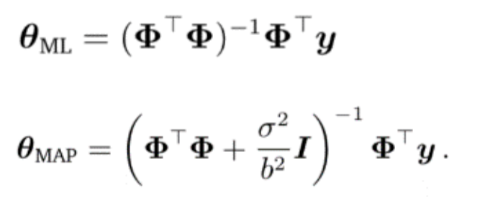
Description automatically generated

Text, letter

Description automatically generated

* if and are independent.
* Change of variables technique: given a random variable *X* and a transformation , we have

Example: if , then

* MLE estimation: usually set the derivative of the negative log likelihood of the variable that we are trying to estimate to 0. Note that if , .
* MAP estimation: maximise such that . The idea is that we are adding a prior term so that overfitting can be reduced. 
* Marginal: (similar for discrete)

## Linear algebra

* Finding eigenvalues of a matrix *A*: set and solve for . Then use to find the eigenvector for each eigenvalue solution .
* If , from above, then *D* is a diagonal matrix of the eigenvalues. Additionally, if is symmetric, then the column vectors in *Q* represent the eigenvectors and form an orthonormal basis ().
* Graphical user interface, text, application

  Description automatically generated   
  - this is usually squared in practice.
* Rayleigh quotient: is strictly inclusively between the smallest and largest eigenvalue of *A*.
* SVD decomposition: to get the SVD of a matrix *A*,   
    
  1. Find and get its eigenvalues, normalising them to 1 as needed. Find their eigenspace and put it into P. Put the eigenvalues into a diagonal matrix *D*. At this stage one should have a form, as is always symmetric and positive semidefinite for any matrix *A*. This should be column-wise, so if you have a pair of eigenvectors , then   
  2. Construct another matrix which is diagonal and is the square root of each of the eigenvalues. So if , . Notice that as is positive semidefinite, the case of a negative square root cannot happen.  
  3. For each diagonal element *i* in Σ and eigenvector in *P*, find each of the elements that comprise *U* by .

4. Get the SVD: .

Note that U isn’t necessarily unique, though a common convention (which is shown in the exercise answers on Scientia) is to order the eigenvalues in descending order (which should be unique if followed).

* Key place: page 158 of the MML book for the gradients.

## Optimisation

* Finding the Lagrangian dual:   
    
  1. Write the Lagrangian (this would be the objective – (constant) constraints)

2. Find the derivative with respect to the objective variable and substitute (if needed)

3. The new objective function in the dual are the terms purely in lambda – the rest go in the constraint. See 7.6 and 7.7 for examples (in the former the derivative is purely in terms of lambda – which means that there is no need to substitute in the Lagrangian).