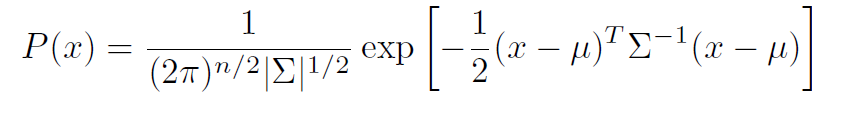
**Writeup on** **Gaussian Markov Networks & Linear Regression on Continuous Variables.**

It is common to use the Linear Regression model to compute dependencies between continuous variables and perform inference. In cases where there are a large number of dependencies and independencies between groups of variables, it may be more convenient to represent the variables as a Bayesian network. Representing this relationship as a network allows to compute joint distributions, conditional distributions and marginal distributions between groups and combinations of variables in the network. In the cases of smaller networks, the inference and joint/marginal distributions can be computed using the Standard form of the Normal distribution. However, for larger networks, it is more convenient to use the Information form to represent Joint distribution, as relationships between observed and latent variables are usually conditional in nature. Using information form makes it easier to carry out inference on a variable or groups of variables.

A Bayesian network is a graph structure representing conditional relationships and independencies between variables having a probability distribution. A Gaussian Bayesian Network represents relationships between a set of normally distributed variables. The relationships between Gaussian variables are always linear.

The multivariate Normal is on continuous vector variable given below



Consider the following Bayesian Network of continuous variables. This is a simple representation of a Linear Regression Equation in graphical form. The network given below, is very simple, however if a more complex network is used, it is possible to appreciate the dependencies and independencies between variables ( for example, X1 is independent of X2).

**Figure 1**

X

X1

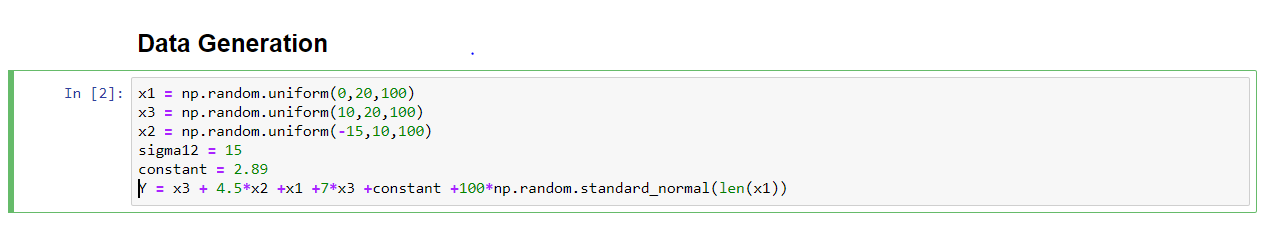
Y

X3

X2

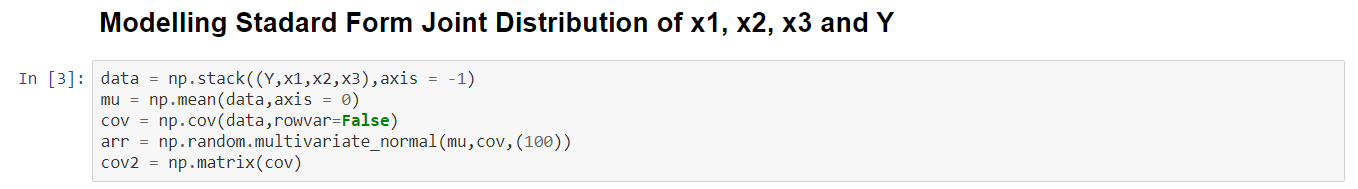
**OR**

y



**JOINT DISTRIBUTION: STANDARD FORM**

The Joint Distribution of this network, where is the joint mean vector and is the joint covariance matrix. The same can be expressed in expanded form as-



**CONDITIONAL DISTRIBUTION: STANDARD FORM**

The Standard equation of Linear Regression is given in vector form.

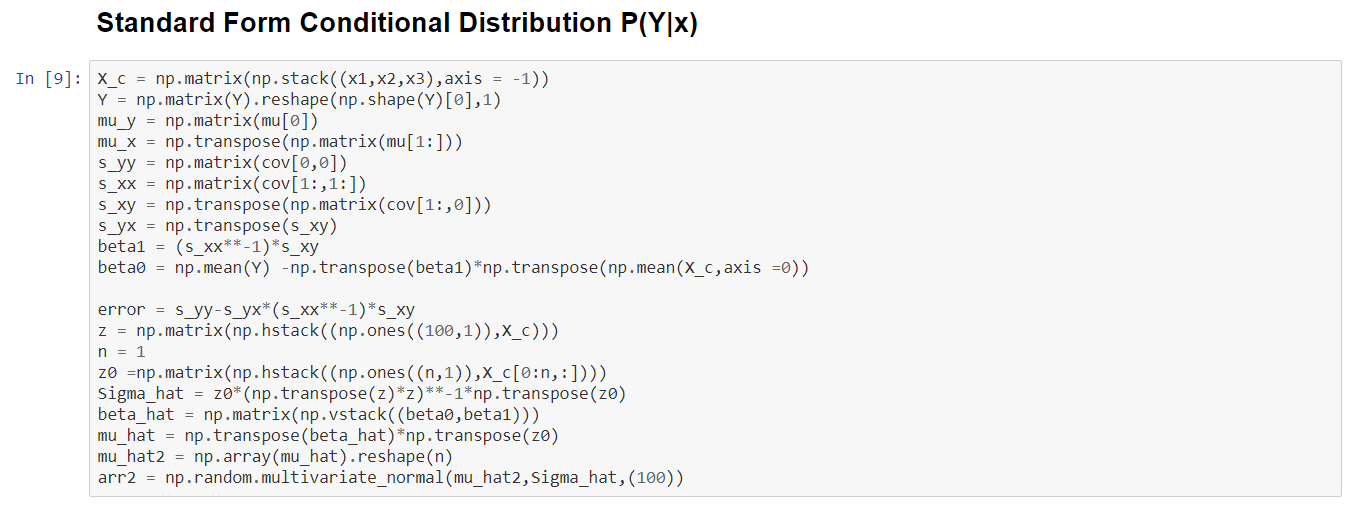
The conditional distribution P(Y|X) models the linear regression equation This is given as -

This can equivalently be represented as

and the scalar

The above-mentioned equations used to predict y for a given point X= Xo vector of points , in which case -

predictive distribution for the point . Given below is the implementation of the conditional distribution in code. A distribution for y|x is computed given a vector .



For larger networks, it is more convenient to calculate joined distributions, conditionals and marginals using the information form of the Gaussian distribution.

**JOINED DISTRIBUTION: INFORMATION/CANONICAL FORM**

The Information Form of the joint distributions, is extremely useful for computing conditional distributions, in graph operations. The Information Matrix itself, is able to represent overall graph structure.

* If an element , then it can be implied that .
* If an element , then it can be inferred that

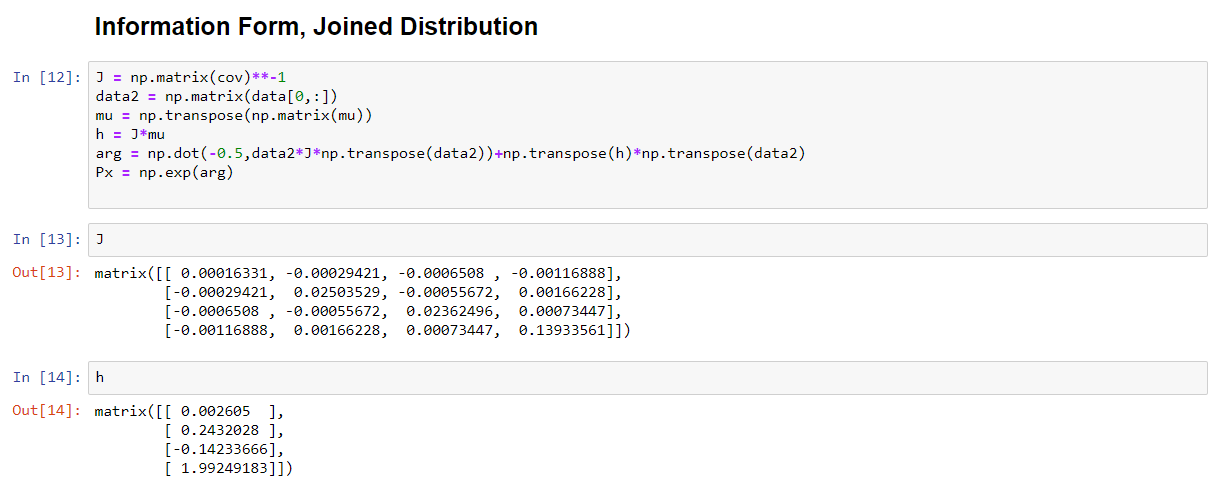
For the problem described above the information matrix of P(x1,x2,x3,y) can be inferred as follows-

The non-zero elements of the represents edges on the graph. The diagonal elements represent the nodes on the graph. The edges , , , are represented as non-zero elements in the precision matrix. Zero values imply that no edges exist between the nodes indicated by the place-holder of this value. Eg if no edge exists between X2 and X3.

The Information form of the gaussian is given below. Let the stacked matrix (Y,X) be represented by S.

S=(Y,X1, X2,X3)

The Joined distribution is implemented in code as follows. The partition function Z has not been computed below.



The conditional distribution P(Y|X) can be computed in the information form by modifying normalizing constants. It is easy to compute the pairwise joint distributions using the information form as shown below- If P(Y,X) is represented by an non- normalized potential the –

=

In the above representation, the potential is the non-normalized form of probability. It must be divided by a Integral of the potentials over all the variables in joint distribution to give a probability value. The Denominators in these formulae are called Partition Functions (represented as Z) and are computationally very intensive.

Computing the Joint distributions for Graph is easy using the information form and may be computed by the following equations-.

For the graph given in Figure 1, the joint distribution can be computed as –

More generally, the joint distribution over a graph is the product of all node potentials X and all edge potentials and given by the equation where is the node potential and is the edge potential between Products of are taken over all nodes and edges respectively. The inference is best performed in these cases by message passing algorithms, which I intend to cover in subsequent writeups.