IEE575 - Applied Stochastic Operations Research Models
Lab 2

Tanmay Khandait
tkhandai@asu.edu

April 14, 2025

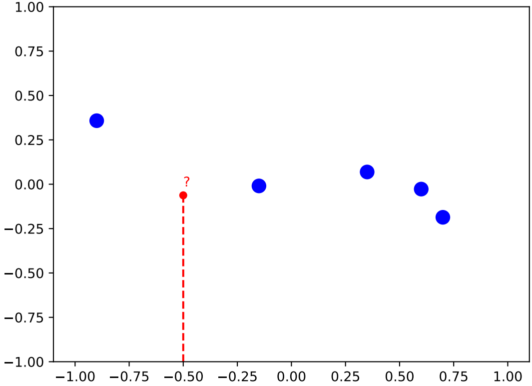
Plan for the next 2 sessions

Today:

- Python, NumPy
- Revisit Univariate and Multivariate Gaussian Distributions.

Motivation

Consider the following regression problem:
What would be the output be at the point in red?



Motivation

Consider the following regression problem:
What would be the output be at the point in red?

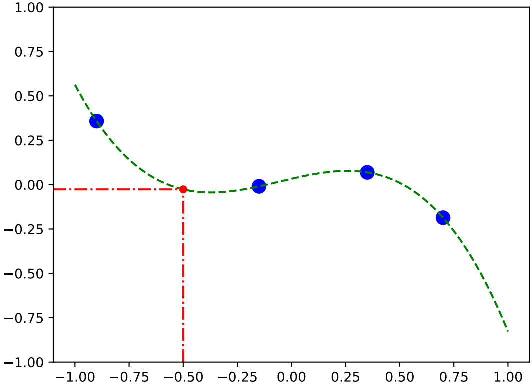


Figure: Here is an illustration of using non-linear regression.

Motivation

Consider the following regression problem:
What would be the output be at the point in red?

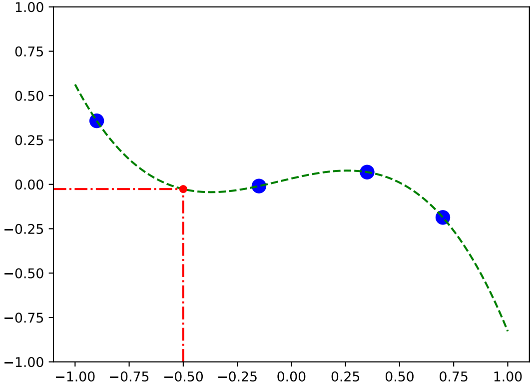


Figure: Here is an illustration of using non-linear regression.

However, in addition to this, we also need uncertainty estimates.

Motivation

Can we do this using what we know about Gaussian Distributions?

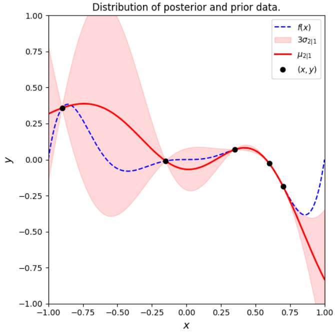


Figure: Here is an illustration of using non-linear regression.

Univariate Gaussian Distribtution

- A random variable X is normally distributed with mean μ and variance σ^2 if it has the probability density function of X as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

.

  - In this expression, you see the squared difference between the variable x and its mean, μ.
  - This value will be minimized when x is equal to μ.
  - The quantity $-\frac{x-\mu^2}{\sigma^2}$ will take its largest value when x is equal to μ or likewise since the exponential function is a monotone function, the normal density takes a maximum value when x is equal to μ.
  - The variance σ^2 defines the spread of the distribution about that maximum. If it is large, then the spread is going to be large, otherwise, if the value is small, then the spread will be small.
- If X is random variable that follows a normal distribution with mean μ and variance σ^2, then we will denote it as $X \sim \mathcal{N}(\mu, \sigma^2)$.

If you are not familiar, play around with this link:
`https://demonstrations.wolfram.com/TheNormalDistribution/`

Multivariate Gaussian Distribtution

- The multivariate normal distribution of a k-dimensional random vector $\boldsymbol{X} = (X_1, X_2, ..., X_k)^T$ is written as $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.
- The probability density function is given as follows:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

  where,
  - $\boldsymbol{\mu} = E[\boldsymbol{X}] = (E[X_1], E[X_2], ..., E[X_k])^T$
  - $\Sigma_{i,j} = E[(X_i - \mu_i)(X_j - \mu_j)] = Cov[X_i, X_j]$

Fun fact: correctly constructed covariance matrices are always symmetric and positive semi-definite. And thus invertible.
But let's look at the covariance matrix in detail.

Covariance Matrix I

Let us look at the covariance matrix in detail: $\Sigma_{i,j} = E[(X_i - \mu_i)(X_j - \mu_j)] = Cov[X_i, X_j]$

$$\Sigma_{k\times k} = \begin{bmatrix} \sigma_1^2 & cov(x_1,x_2)^2 & \dots & cov(x_1,x_{k-1}) & cov(x_1,x_k) \\ cov(x_2,x_1)^2 & \sigma_2^2 & \dots & cov(x_2,x_{k-1}) & cov(x_2,x_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ cov(x_{k-1},x_1)^2 & cov(x_{k-1},x_2)^2 & \dots & \sigma_{k-1}^2 & cov(x_{k-1},x_k)^2 \\ cov(x_k,x_1)^2 & cov(x_k,x_2)^2 & \dots & cov(x_k,x_{k-1})^2 & \sigma_k^2 \end{bmatrix}$$

- The diagonal elements of the matrix contain the variances of the variables.
- The off-diagonal elements contain the covariance between all possible pairs of variables.

Question: What happens when the off-diagonal elements are 0? What does it mean?

Let us look at a bivariate Gaussian Distribution for different values of the covariance matrix.

Covariance Matrix II

$$\boldsymbol{\mu}_{2\times 1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma_{2\times 2} = \begin{bmatrix} 1 & 0.0 \\ 0.0 & 1 \end{bmatrix}$$

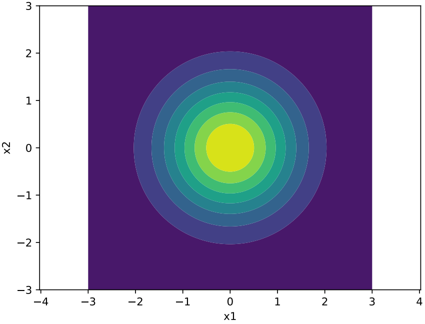


Figure: Distribution of Points

Covariance Matrix III

$$\boldsymbol{\mu}_{2\times 1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma_{2\times 2} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}$$

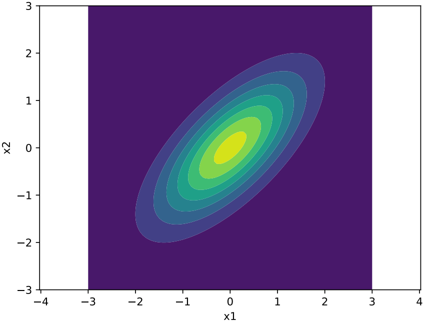


Figure: Distribution of Points

Covariance Matrix IV

$$\boldsymbol{\mu}_{2\times 1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma_{2\times 2} = \begin{bmatrix} 1 & -0.7 \\ -0.7 & 1 \end{bmatrix}$$

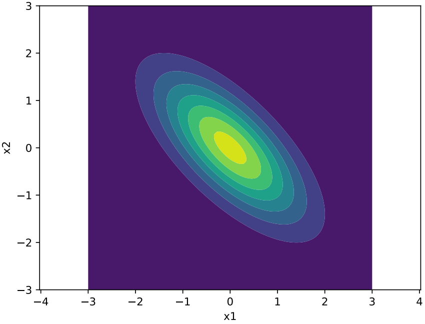


Figure: Distribution of Points

Marginalization

- Given a multivariate distribution, can we compute the pdf of a single variable? - Yes
  $f(X_1) = \int f(X_1, X_2) dX_2$
- Every random variable $X_i \in \boldsymbol{X}$ has the following distribution: $X_i \sim \mathcal{N}(\mu_i, \Sigma_{i,i})$

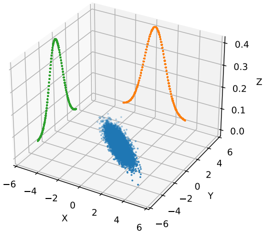


Figure: Covariance Matrix

$$\Sigma_{2\times 2} = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$$

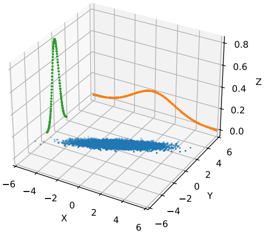


Figure: Covariance Matrix

$$\Sigma_{2\times 2} = \begin{bmatrix} 2 & 0.9 \\ 0.9 & 0.5 \end{bmatrix}$$

Conditioning

If some random variables from the set were fixed, what is the PDF?
Or formally, $f(X_i | X_j = x) = ?$

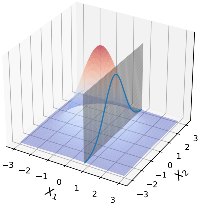


Figure: Covariance Matrix

$$\Sigma_{2\times 2} = \begin{bmatrix} 1 & 0.0 \\ 0.0 & 1 \end{bmatrix}$$

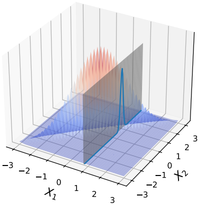


Figure: Covariance Matrix

$$\Sigma_{2\times 2} = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}$$

- In fact, it is a Gaussian. Or formally, $X_i | (X_j = x) \sim \mathcal{N}(\mu_*, \Sigma_*)$.
- In a bivariate case, $f(X_2 | X_1 = x_1) \propto \exp(-\frac{1}{2}(x_2 - \mu_*)\Sigma_*^{-1}(x_2 - \mu_*))$.
- We will see what μ_* and Σ_* looks like in a few slides.

Conditioning

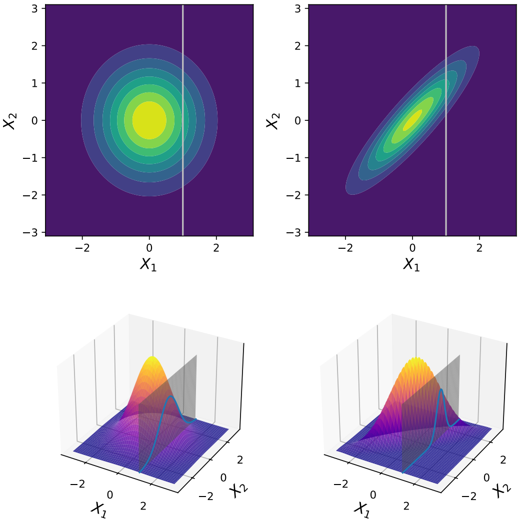
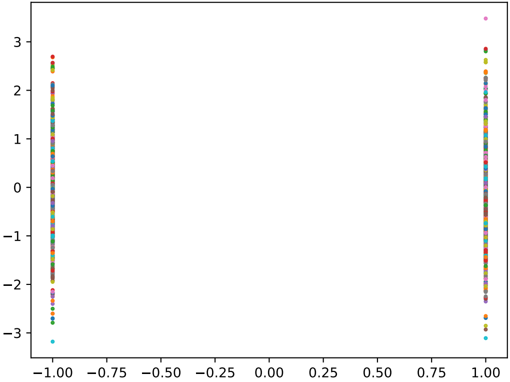


Figure: Detailed View of Conditioning

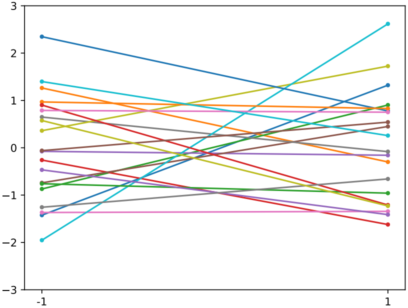
Intuition

- Let us sample two random variables $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$.
- Next, we can plot multiple independent Gaussian in the coordinates. For example, put vector X_1 at $x = -1$ and another vector X_2 at $x = 1$.



Intuition

- Let's connect points of X_1 and X_2 by lines. For now, we only generate 20 random points, and then join them up as 10 lines. Keep in mind, that these randomly generated 10 points are Gaussian.



Going back to think about regression. These lines look like functions for each pair of points. On the other hand, the plot also looks like we are sampling the region with 20 linear functions even though there are only two points on each line. In the sampling perspective, the domain is our region of interest, i.e. the specific region we do our regression.

- Let's connect points of X_1, X_2, and X_3 by lines. For now, we only generate 10 random points, and then join them up as 10 lines. Keep in mind, that these randomly generated 10 points are Gaussian.

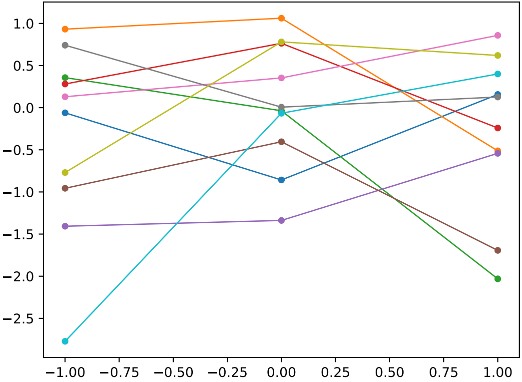


Figure: X_1 will be at -1, X_2 will be at 0, and X_3 will be at 1

- Let's connect points of X_1, X_2, X_3, X_4, and X_5 by lines.

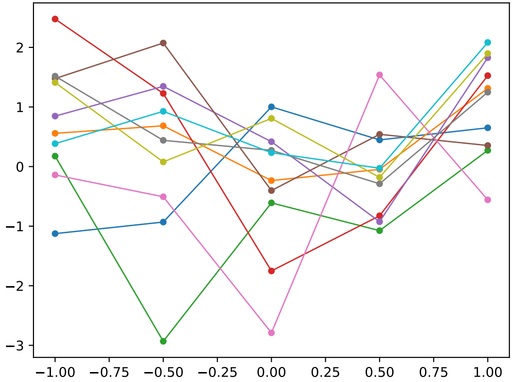


Figure: X_1 will be at -1, X_2 will be at -0.5, X_3 will be at 0, X_4 will be at 0.5, and X_5 will be at 1

Intuition

This sampling looks even more clear if we generate more independent Gaussian and connecting points in order by lines.

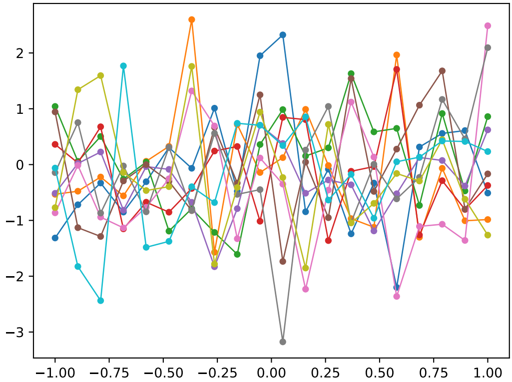


Figure: X_1 will be at -1, X_2 will be at -0.89, X_3 will be at -0.78, $\ldots$, X_{19} will be at 0.89, and X_{20} will be at 1

What are the problems?

Question: We don't have a way of correlating these independent distributions. How do we solve this?

Use Multivariate Gaussian Distributions.

Question: How do we consider the points that we already have as prior?

Use Conditioning
Let's solve these two problems first!

Revisiting Conditioning - Mathematically

So what exactly is the mean and variance when there is a conditional multivariate distribution?

$$\boldsymbol{X} = (X_1, X_2) \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}, \begin{bmatrix}A & B\\ B^T & C\end{bmatrix}\right)$$

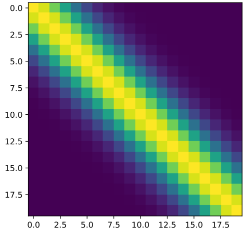
$$X_1|X_2 = \frac{X_1, X_2}{X_2}$$

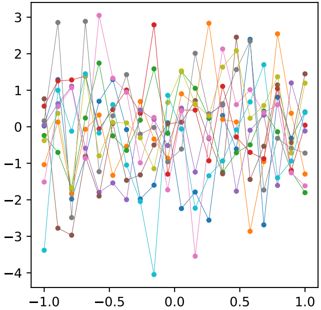
$$X_1|(X_2 = \boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{a} + BC^{-1}(\boldsymbol{x} - b), A - BC^{-1}B^T)$$

Do you notice the predictive mean is linear in data?
Do you notice that the predictive uncertainty is prior uncertainty minus the reduction in uncertainty?

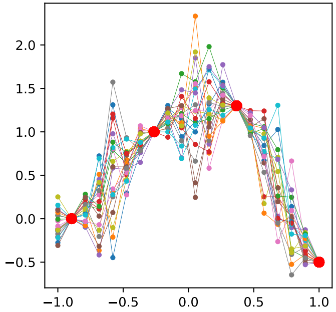
Visualizing Multivariate Distributions with Conditioning

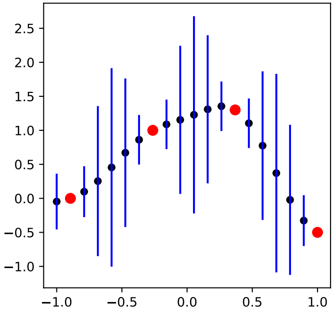
We have $\boldsymbol{X}$ which is drawn from a Multivariate Distribution with 20 components. In this case, X_1 will be at -1, X_2 will be at 0.89, X_3 will be at -0.78, ..., X_{19} will be at 0.89, and X_{20} will be at 1.





Let us fix $X_1(x=-0.89), X_7(x=-0.26), X_{13}(x=0.37), X_{19}(x=1.0)$.





Next Class

We are almost ready to implement Gaussian Process Regressors. We will:

- Implement our own gaussian process regressor
- Scikit-Learn
- GPy

Let's dive into code.