Topic :⇒ Gaussian Process Regression Model.

Multivariate Gaussian Distribution :⇒ Multivariate Gaussian Distribution ~~(MVN)~~ (MVGD) is what Gaussian Processes are based upon. It is defined by mean vector and Covariance matrix.

$$X \sim \mathcal{N}(\mu, \Sigma)$$

It possesses two properties :

i) Marginalisation :⇒ With Joint Gaussian distribution, it can be written as

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$

marginalised Y with resulting X can be written as,

$$p(X) = \mathcal{N}(\mu_X, \Sigma_{XX})$$

ii) Conditioning :⇒

$$p(X \mid Y) = \mathcal{N}(\mu'_X, \Sigma'_{XX})$$

where $\mu' = \mu_X + \Sigma_{XY} \Sigma_{XY}^{-1} (Y - \mu_Y)$

$$\Sigma'_{XX} = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX}$$

Now for Gaussian Process, in the same way we define GP as

$$f(x) = GP(m(x), K(X, X'))$$

where $m(x)$ = mean function.
$K(X, X')$ = Kernel function .. (discussed later)
$X$ = vector
$f(x)$ = real-valued function.

Kernels : These are functions that are used to generate covariance matrix.

In this case our kernel function has 2 parameters :

$\sigma$ : scale of variance

$l$ : influence of pts to neighbouring points.

$$k(x_1, x_2) = \sigma^2 \cdot \exp\left( \frac{(x_i - x_j)^2}{2l^2} \right)$$

——— x ——— x ——— x ———

For Given test data $x'$, we have to predict $Y'$

Hence,

$$\begin{pmatrix} Y \\ Y' \end{pmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_Y \\ \mu_{Y'} \end{bmatrix}, \begin{pmatrix} K & K' \\ K'^T & K'' \end{pmatrix} \right)$$

$K$ = kernel $(Y, Y)$

$K'$ = kernel $(Y, Y')$

$K''$ = kernel $(Y', Y')$

So,

$$p(Y' \mid X', X, Y) \sim \mathcal{N}(\mu_0', \Sigma_0')$$

where $\mu_0' = \mu' + K'^T K^{-1} (X - \mu)$

$$\Sigma' = K'' - K'^T K^{-1} K'$$

Since, we have to perform matrix inversion while trying to find updated covariance. In practical, matrix inversions are avoided due to reason:

i) Numerically Unstable.

ii) Computationally heavy :=) Calculating inverse takes $O(N^3)$ time complexity

Since K is semi-positive definite matrix, it can be decomposed ~~into~~ using cholesky decomposition.

$$K = LL^T$$

where $L$ = Lower - Triangular Matrix with entries $\in \mathbb{R} > 0$.

Now, we know that

$$\mu_0 = \mu' + K'^T K^{-1} (x - \mu)$$

let $\alpha = K^{-1} (x - \mu)$

$$\alpha = (LL^T)^{-1} (x - \mu)$$

$$\alpha = L^{-T} L^{-1} (x - \mu)$$

let $\gamma = L^{-1} (x - \mu)$

$$L\gamma = (x - \mu)$$

$$\gamma = L/(x - \mu)$$ if $AW = B$ then $W = A/B$.

then $\alpha = L^{-T} \gamma$

$$\alpha = L^T / (L / (x - \mu))$$

So

$$\mu_0 = \mu' + K'^T \alpha.$$

Similarly for $\Sigma_0$

let $V = L / K'$

$$V = L^{-1} K'$$

$$V^T V = K'^T L^{-T} L^{-1} K'$$

Hence

$$\Sigma_0 = K'' - V^T V$$