(3.) Mathematics .

Given $(x, f(x))$ to find $f(x)$ that passes through all the point.

Normally we find ~~a~~ function ~~w~~ parameters which fits our data well. and then define the function with the parameters.

We have data D we find $(w) \to$ paramth then we predict $y = w^T x$.

In GP instead we try to find the ~~function directly witho~~ prediction directly without limiting to one function.

So we find

$$\boxed{P(y_*|x_*, D)} = \int_w P(y_*|x_*, w)\, P(w|D)\, dw$$

We assume that our data is continuous and follows gaussian distribution.

So we $P(y_*|x_*, w) \to$ Gaussian and $P(w) \to$ Gaussian, $P(D|w) \to$ Gaussian

$$P(w|D) = \frac{P(D|w)\, P(w)}{Z}$$

and hence $P(w|D)$ will be gaussian.

~~theys~~ Hence $P(y_*|x_*, D)$ will be gaussian.

Now we know $P(y_*|x_*, D)$ is gaussian so we assume our data with both test and training data to be gaussian.

~~P(y~~

$$P([y_1, y_2 \dots y_n, y_*] \,|\, [x_1, x_2 \dots x_n, x_*]) \sim N(\mu, \Sigma)$$

Now we can normalise the data to make the mean 0 hence

$$P\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ y_* \end{bmatrix} \,\middle|\, [x_1 \dots x_n], x^*\right) \sim N(0, \Sigma)$$

where Σ is the covariance matrix the ~~too~~ In linear case ~~th~~ to find covariance matrix we have to do Cholesky decomposition of the variance matrix. Cholesky decomposition decomposes the matrix into $V^T d$ ~~T~~ V which is kind of taking square root.

~~$X \to V V^T$~~ $X \to V^T V$

X can be decomposed into $V^T V$.

Now
X needs to be positive semi-definite to do Cholesky decomposition because

Now $X \to V^T V$

$$y^T X y \to y^T V^T V y$$

$$y^T X y \to (Vy)^T V y$$

$$y^T X y \to \|Vy\|^2$$

$$y^T X y \geq 0$$

Hence X is positive semi definite.

the covariance matrix ~~can be~~ is a kernel
because it is positive semi definite. So
We can define the covariance matrix by a kernel
function. We can use RBF kernel function as
it is infinite dimensional kernel.

RBF kernel $= k(x, z) = e^{-\frac{\|x-z\|^2}{2\sigma^2}}$ σ is a free parameter.

So far we have

$$p\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ y^* \end{bmatrix} \middle| \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ x^* \end{bmatrix}\right) = \mathcal{N}(0, \Sigma)$$

where
$\Sigma = K =$
$K_{ij} = k(x_i, x_j)$

~~We now need to~~
~~w~~ But we know $[y_1, y_2 \ldots y_n]$ given in our data and we need to find y^* at our ~~training~~ testin. point.

$$p(y^* \mid y_1, y_2 \ldots y_n, x_1 \ldots x_n, x^*).$$

Pf

By conditioning

$$P(y^* | (y_1, y_2 \ldots y_n\, x_1, x_2 \ldots x_n x^*) = N(\mu_{y^*|(y_1 \ldots x_n x^*)}, \Sigma_{y^*|(y_1 \ldots x_n x^*)})$$

By conditioning identity.

$$\mu_{y^*|(y_1 \ldots x_n x^*)} = K_*^T K^{-1} y$$

where K_* is the ~~kern~~ covariance matrix b/w x_i's and x Training points and Testing points.

K is the covariance matrix/kernel matrix of Training dat.

y is our Training data.

and

$$\Sigma_{y^*|(y_1 \ldots y_n x_1 \ldots x_n x^*)} = \cancel{K_*^T K^{-1} K_*}\ K_{**} - K^* K^{-1} K^*$$

where K_{**} is the covariance matrix of of Testing point.

This completes the GP model.

We can define how well our model extrapolates the extrapolated points, by the variance of the extrapolated points, if the variance is very large ~~to~~ then we say that it is poorly extrapolated.

Yes we need to optimize it, we can optimize by choosing the right parameter of the ~~kernel~~ RBF kernel.
