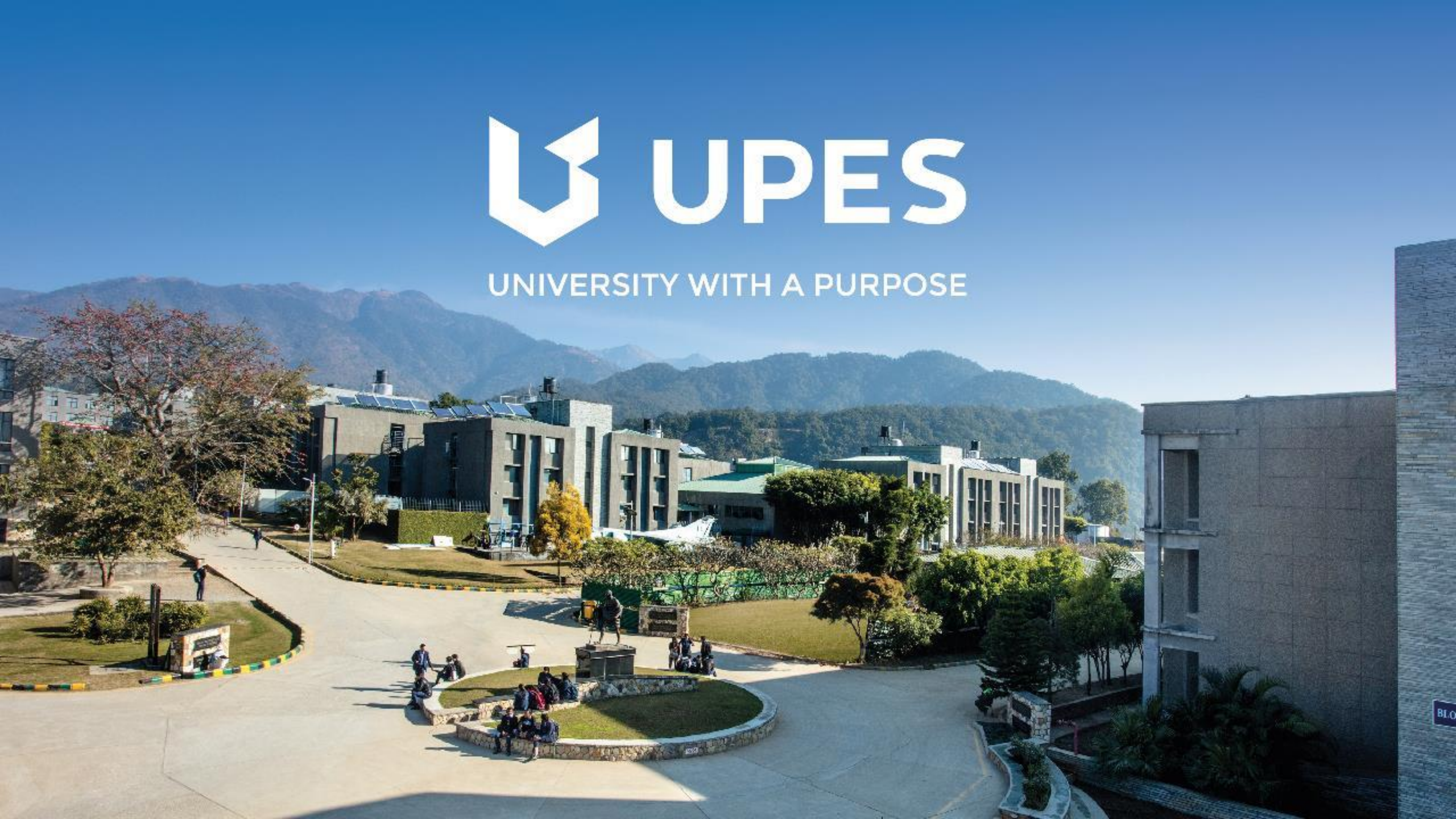
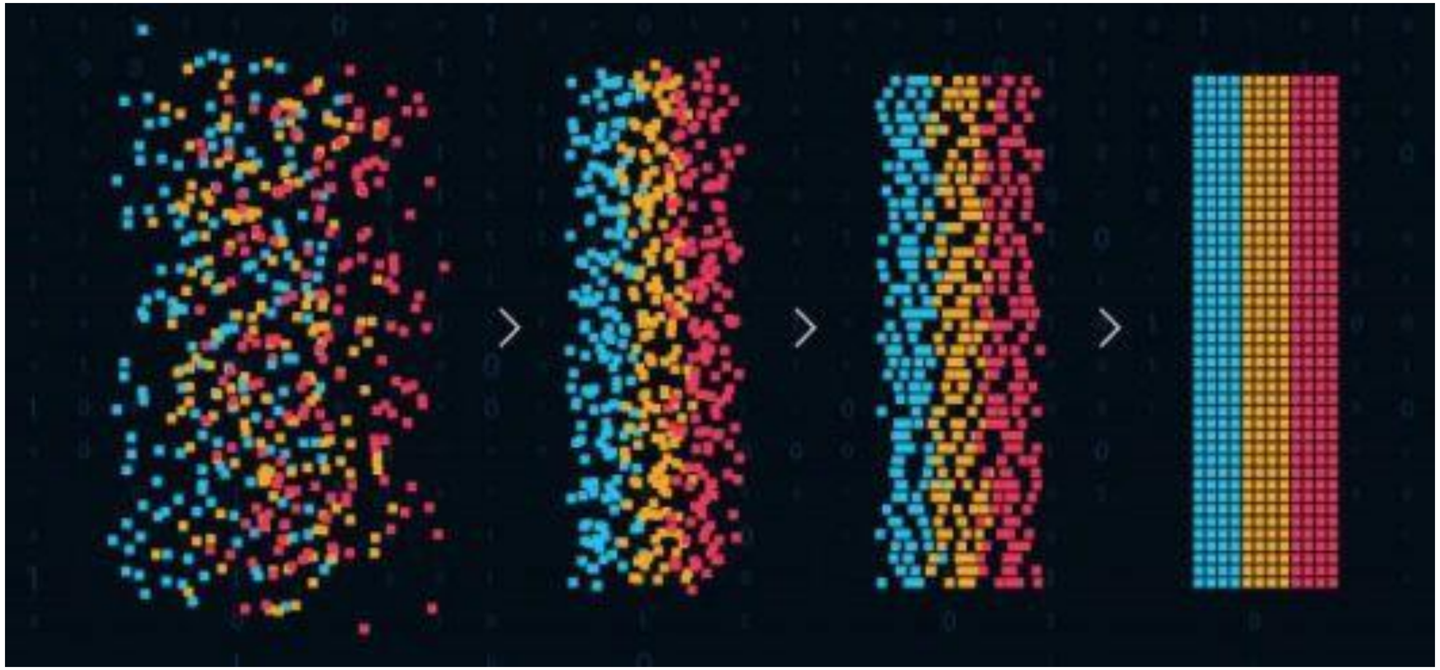

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML

Dr Gopal Singh Phartiyal

18/10/2021

Recap: Linear Regression Models

Goal: Find **w?**

- Why linear model?
- Simple linear regression
- Basis functions
- Solving for w using maximum likelihood and least squares
- For multiple output
- Regularize the model (different regularizers)
- Managing over-fitting via the concept of bias-variance decomposition

So which model to choose?

Remember *model selection* and *cross-validation*

Recap: Bayesian Linear Regression

- Posterior distribution $\boxed{p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)}$

- With a specific prior $\boxed{p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})}$
- Mean and covariance of posterior **w**

$$\boxed{\begin{aligned} \mathbf{m}_N &= \beta \mathbf{S}_N \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_N^{-1} &= \alpha \mathbf{I} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}. \end{aligned}}$$

Example

$$y(x, \mathbf{w}) = w_0 + w_1 x.$$

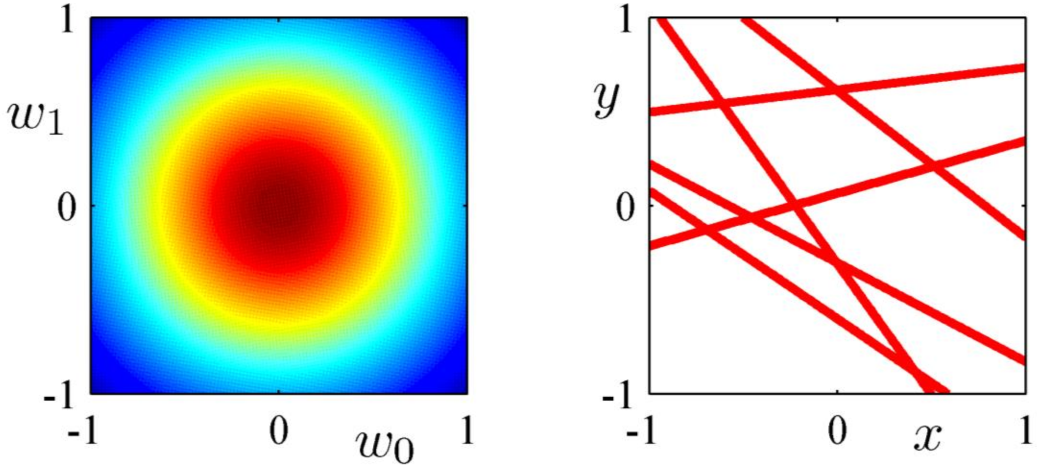
X = U(x| (-1,1), 20 observations

$W_0 = 0.3, \quad w_1 = 0.5$

Noise = Gaussian (sigma = 0.2)

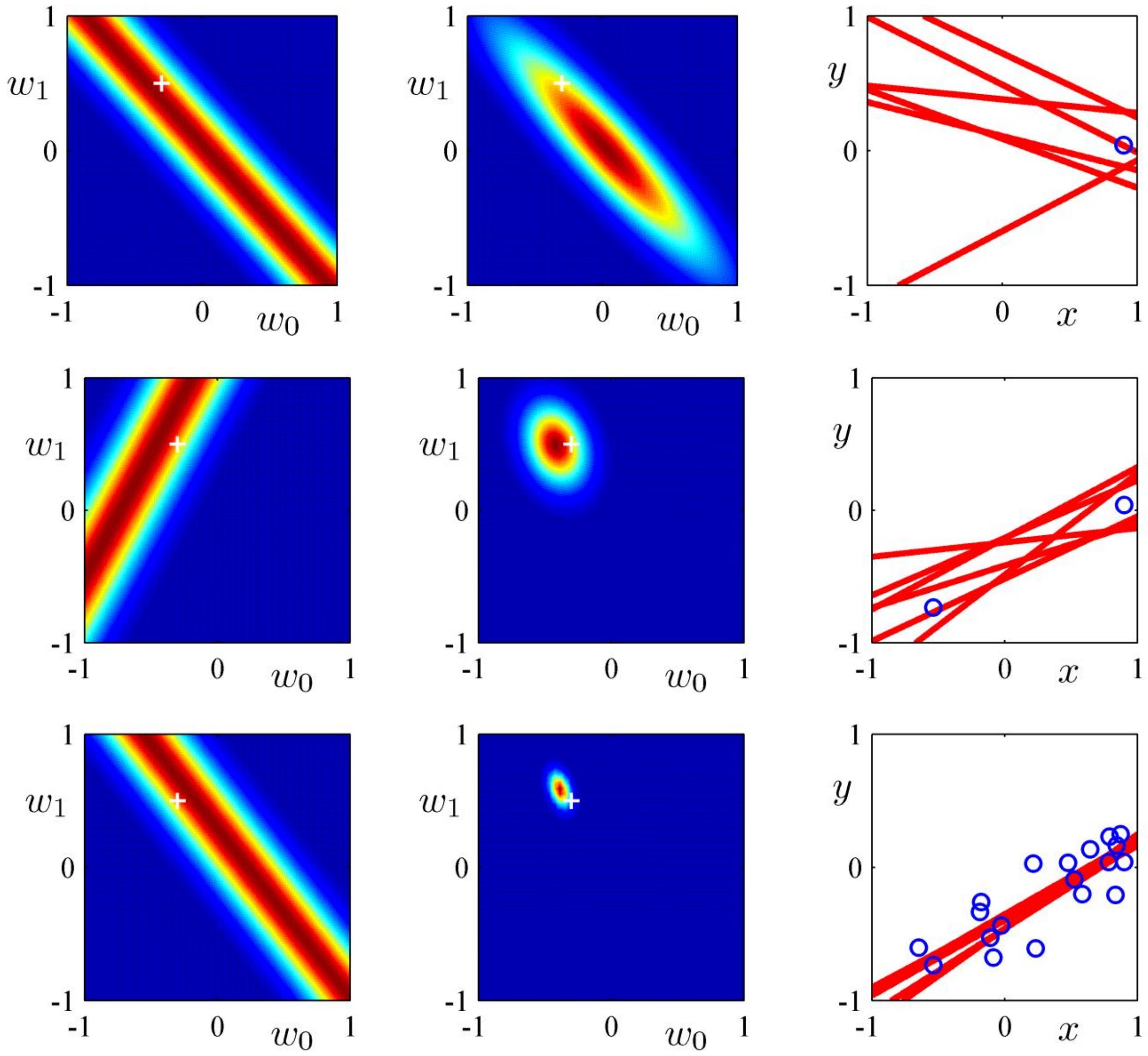
Alpha = 2 for prior

Prior | Data Space



0 data points observed

Likelihood | Posterior | Data Space



Bayesian Linear Regression: Predictive Distribution

- Predict t for new values of x by integrating over w:

$$
\begin{aligned}
p(t|\mathbf{t},\alpha,\beta) &= \int \underbrace{p(t|\mathbf{w},\beta)}_{\text{Conditional}}\overbrace{p(\mathbf{w}|\mathbf{t},\alpha,\beta)}^{\text{Posterior}}\,\mathrm{d}\mathbf{w} \\
&= \mathcal{N}(t|\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}),\sigma_N^2(\mathbf{x}))
\end{aligned}
$$

where

noise on the data

$$
\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{S}_N\boldsymbol{\phi}(\mathbf{x}).
$$

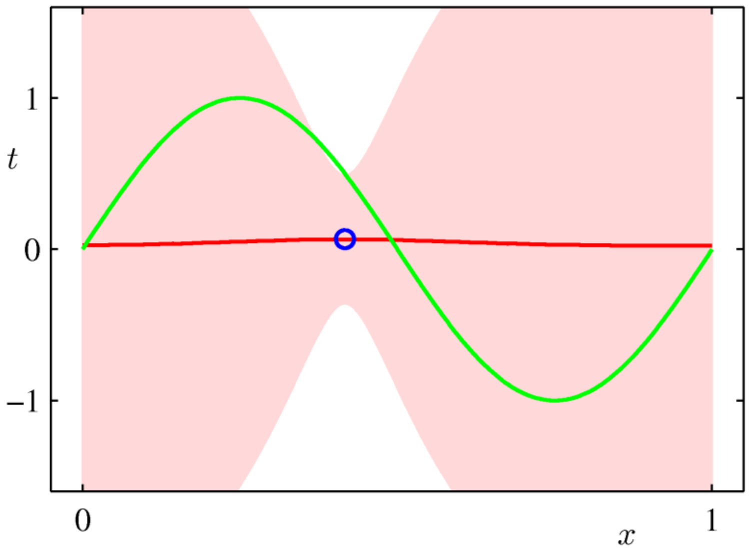
uncertainty associated with the parameters **w**.

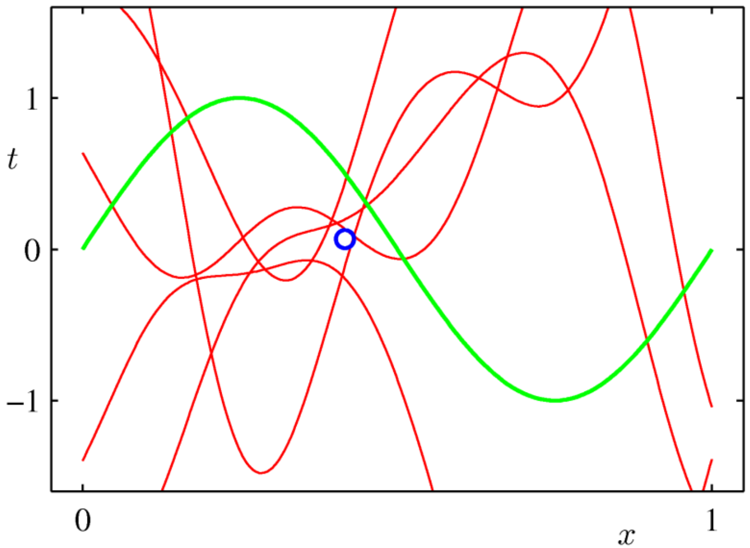
Bayesian Linear Regression: Predictive Distribution

- Example: Sinusoidal data, 9 Gaussian basis functions, 1 data point

$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

Model : Comprising a linear combination of Gaussian basis functions



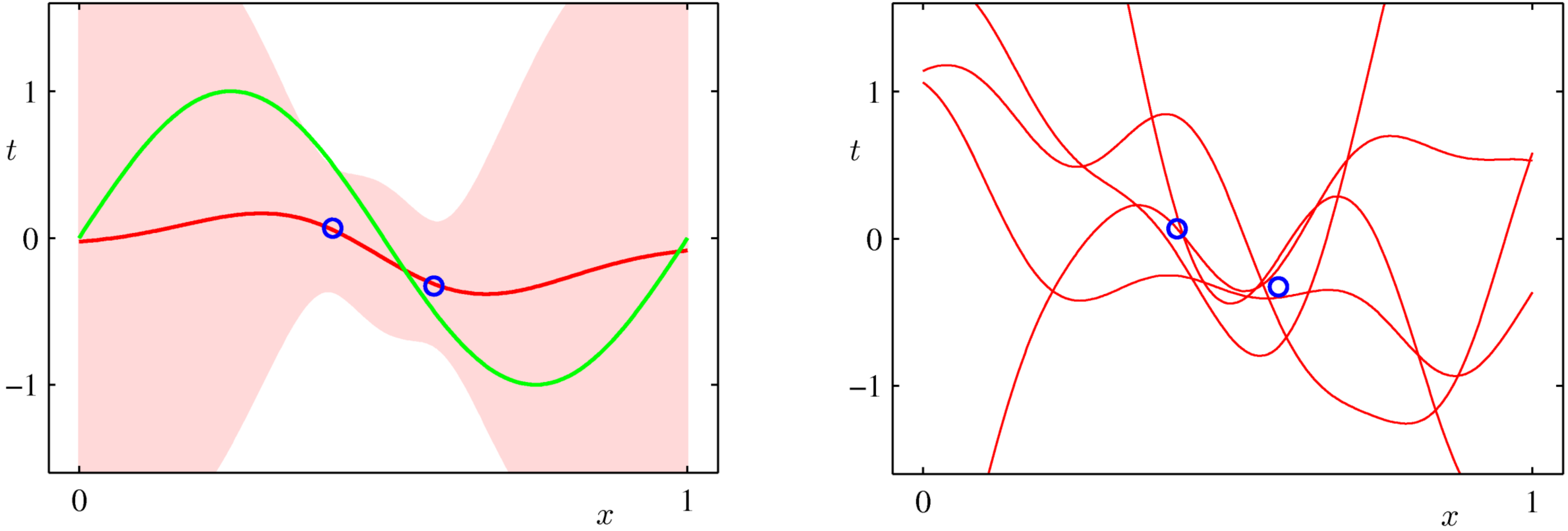


Red curve: Mean of the Gaussian predictive distribution

Shaded region: one standard deviation either side of the mean

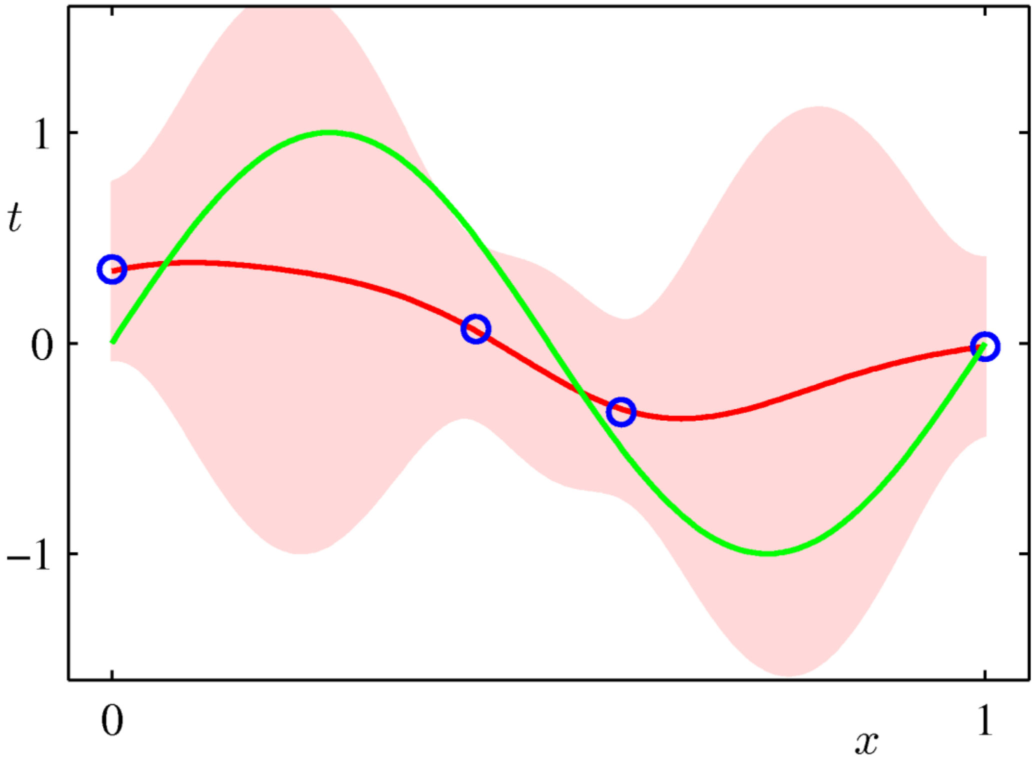
Bayesian Linear Regression: Predictive Distribution

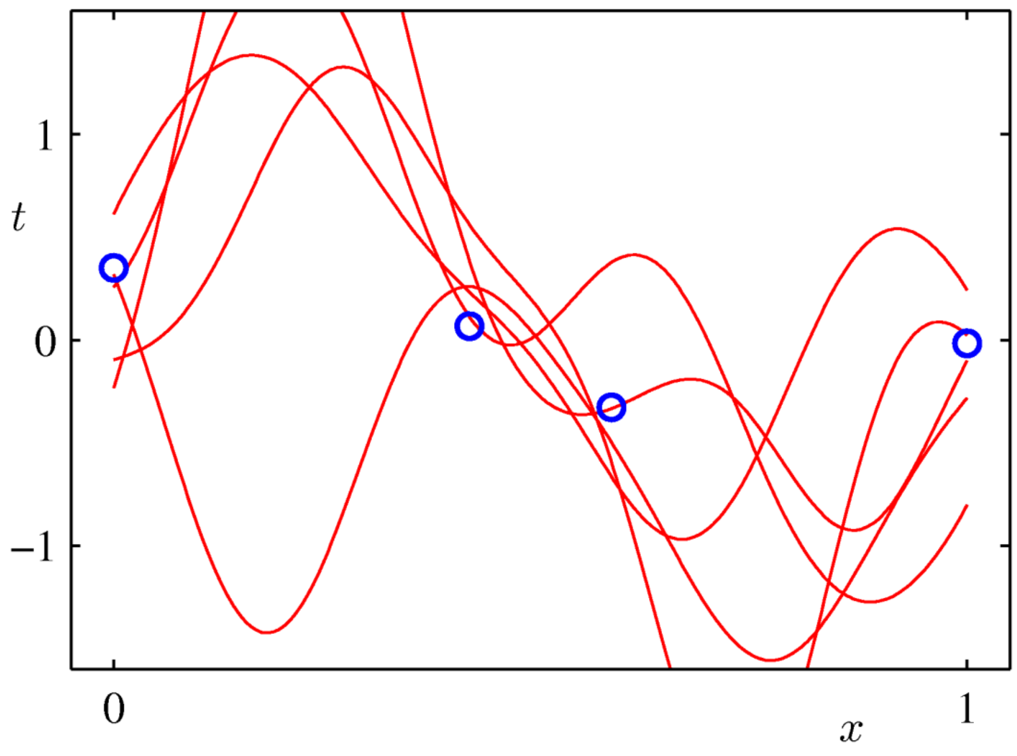
- Example: Sinusoidal data, 9 Gaussian basis functions, 2 data points



Bayesian Linear Regression: Predictive Distribution

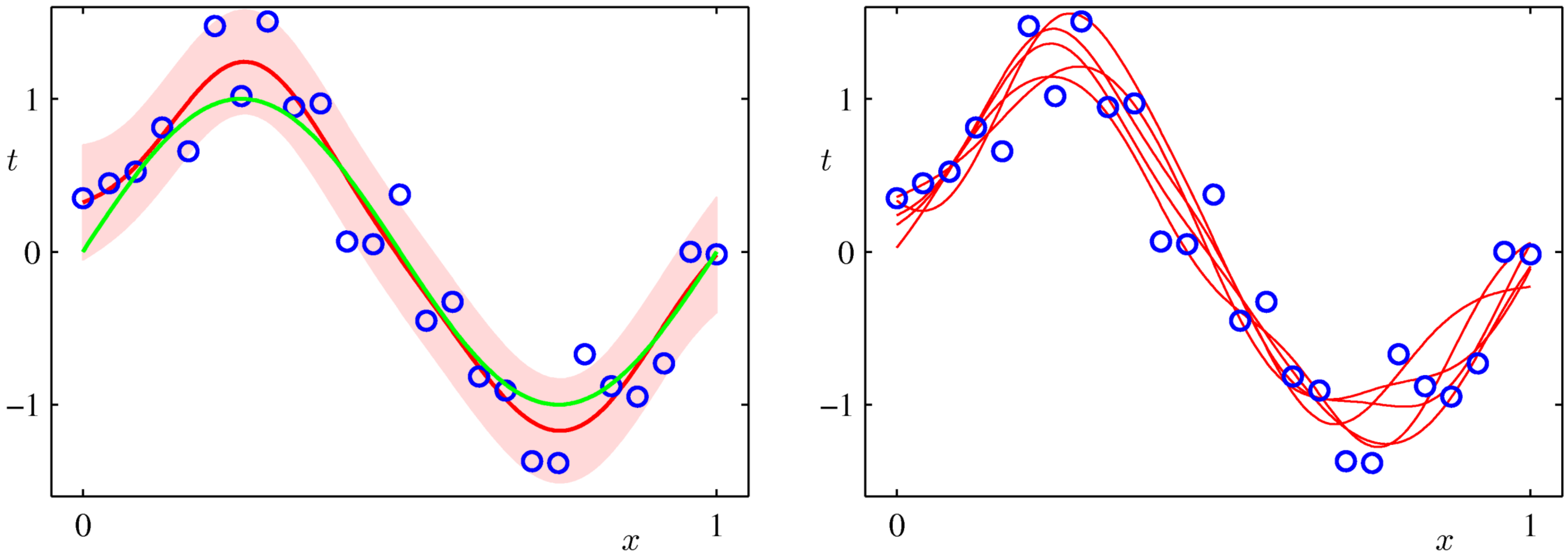
- Example: Sinusoidal data, 9 Gaussian basis functions, 4 data points





Bayesian Linear Regression: Predictive Distribution

- Example: Sinusoidal data, 9 Gaussian basis functions, 25 data points



Equivalent Kernel

- Remember

$$\boxed{y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \ldots}$$

$$\boxed{y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})}$$

- The predictive mean can be written

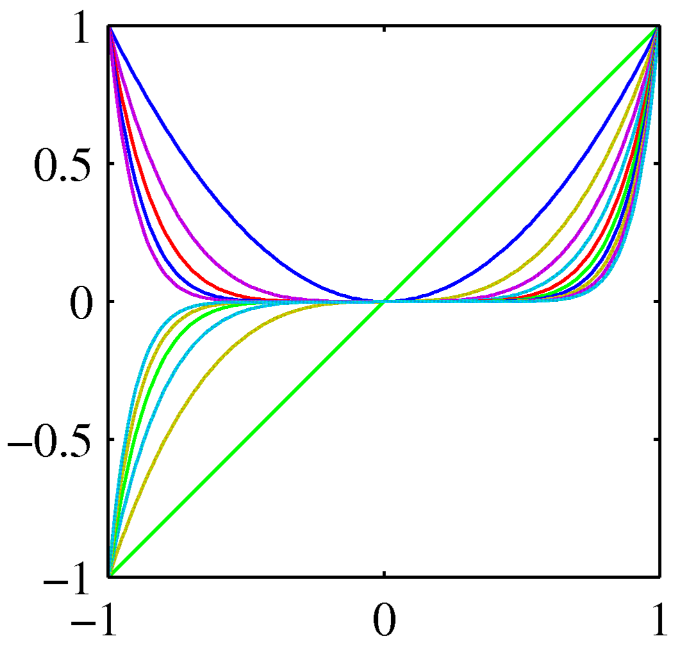
$$\begin{aligned} \mathbf{m}_N &= \beta \mathbf{S}_N \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_N^{-1} &= \alpha \mathbf{I} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}. \end{aligned}$$

$$\begin{aligned} y(\mathbf{x}, \mathbf{m}_N) &= \mathbf{m}_N^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}) = \beta \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} \\ &= \sum_{n=1}^{N} \underbrace{\beta \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}_n)}\, t_n \\ &= \sum_{n=1}^{N} \underbrace{k(\mathbf{x}, \mathbf{x}_n)}\, t_n. \end{aligned}$$

Equivalent kernel or *smoother matrix.*

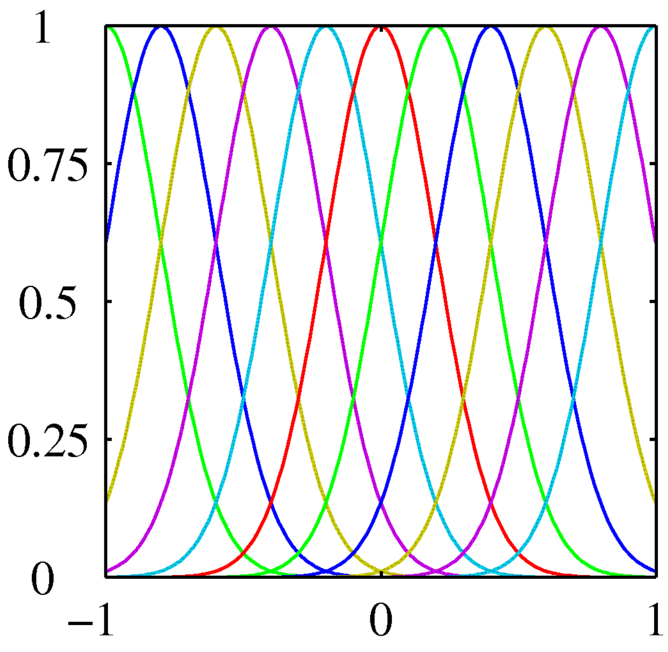
- This is a weighted sum of the training data target values, t_n.

Basis Functions



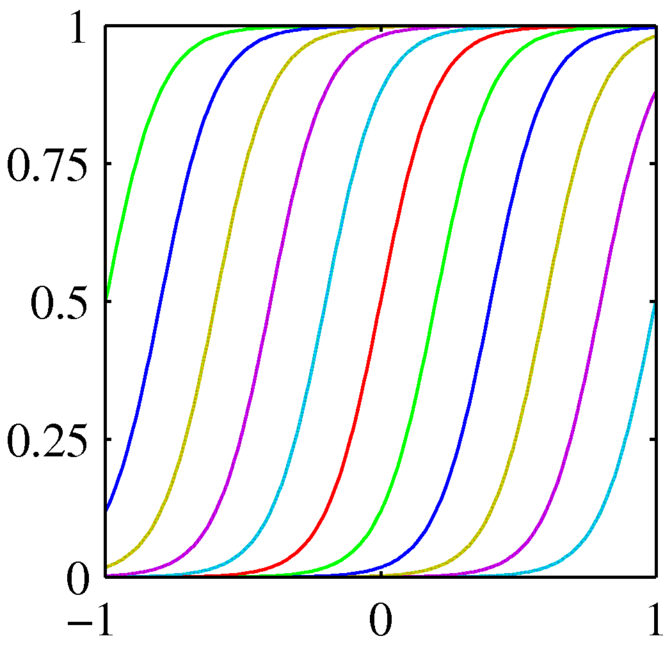
$$\phi_j(x) = x^j.$$

Polynomial basis functions



$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

Gaussian basis functions



$$\phi_j(x) = \sigma\left(\frac{x-\mu_j}{s}\right)$$

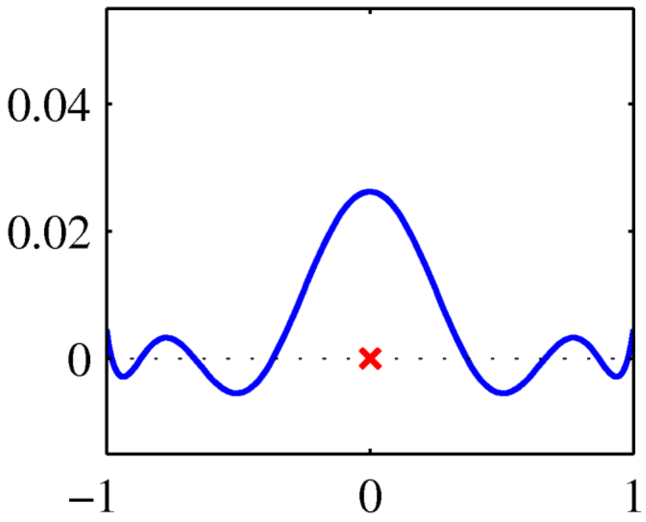
where $\sigma(a) = \dfrac{1}{1+\exp(-a)}.$

Sigmoidal basis functions

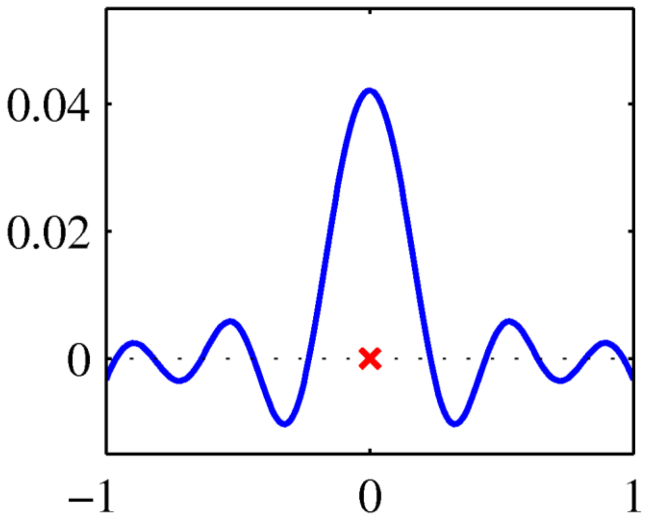
Equivalent Kernel

- Non-local basis functions have local equivalent kernels:

$$k(\mathbf{x}, \mathbf{x}') = \beta \phi(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \phi(\mathbf{x}')$$

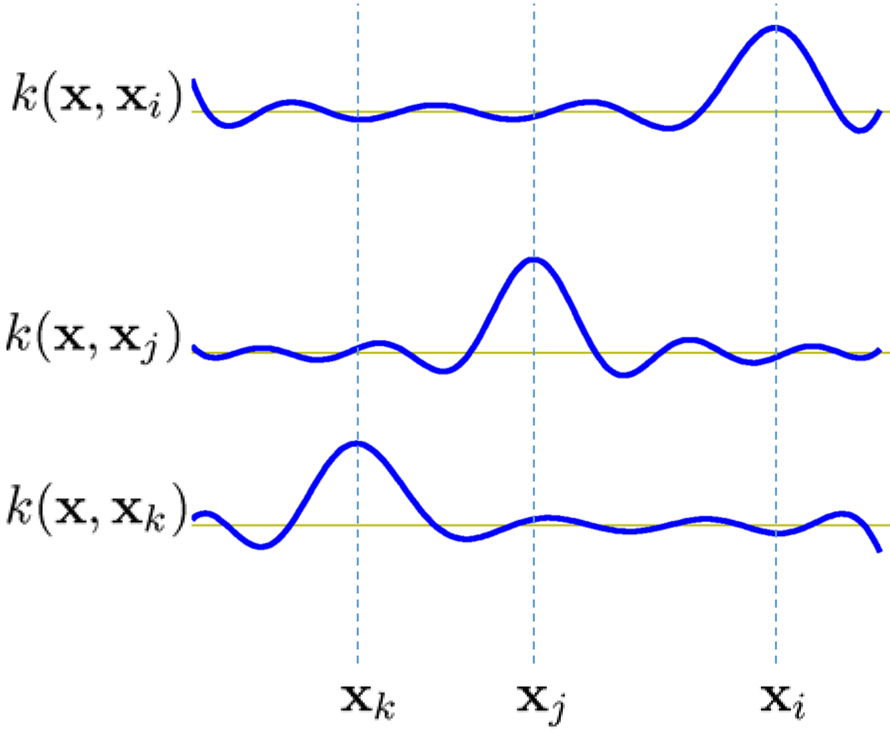


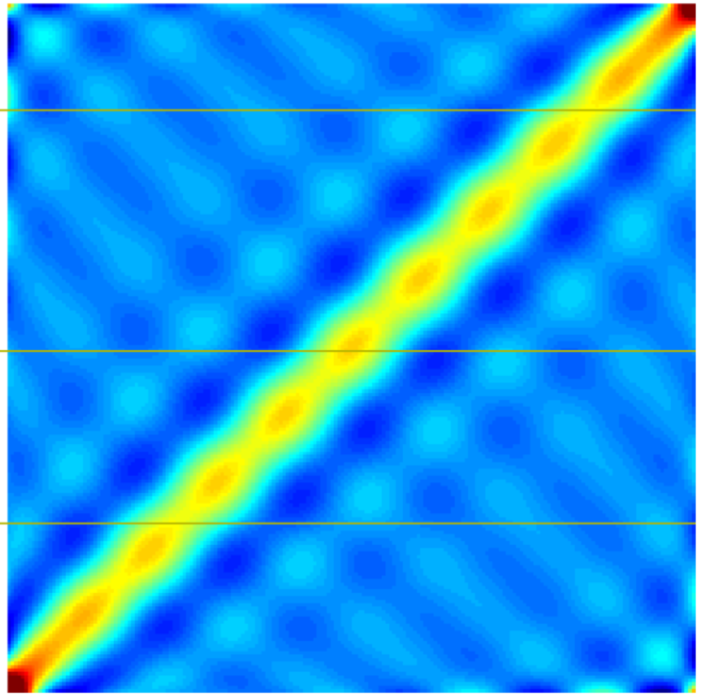
Polynomial



Sigmoidal

Equivalent Kernel





Weight of t_n depends on distance between x and x_n; nearby x_n carry more weight.

Thank You

UPES

UNIVERSITY WITH A PURPOSE