University of BRISTOL

Machine Learning

Additional lecture

Carl Henrik Ek - carlhenrik.ek@bristol.ac.uk

October 17, 2017

http://www.carlhenrik.com

Questions

Q1

Can you do practical questions on each topic?

Q2

$$p(\mathbf{X}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- does this mean the we assume x to be independent?

Q3

$$\mathsf{x}_i = [x_i, 1]^{\mathrm{T}} \quad \mathsf{W} = [w_1, w_0] \quad y \in \mathbb{R}$$

- in the linear regression what is the dimensionality of the variables

Errors

404

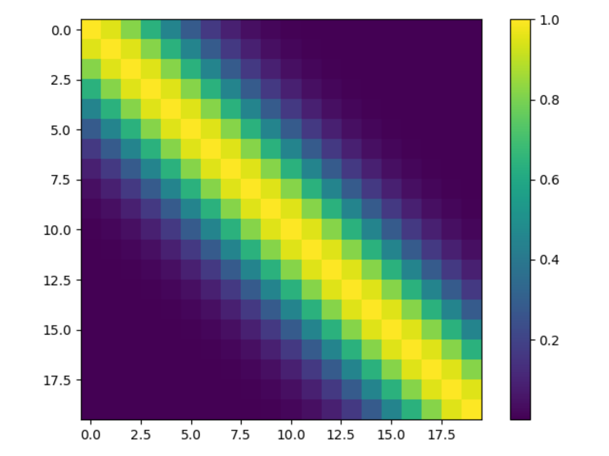
Gaussian Processes

Kernel parameters

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 e^{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)}{l^2}}$$

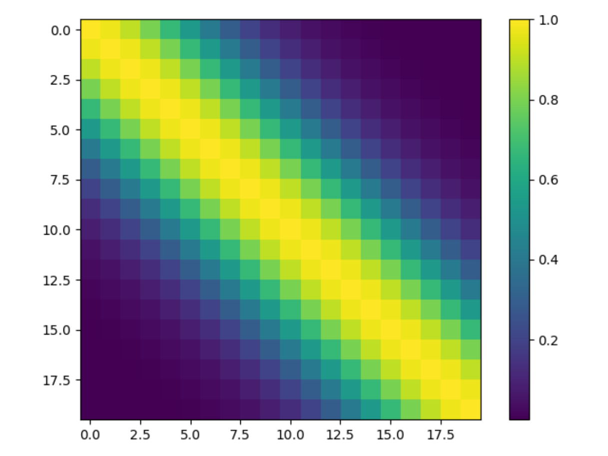
- This co-variance has two parameters, how do they effect the prior, i.e. what assumptions do the encode?

Kernel Parameters



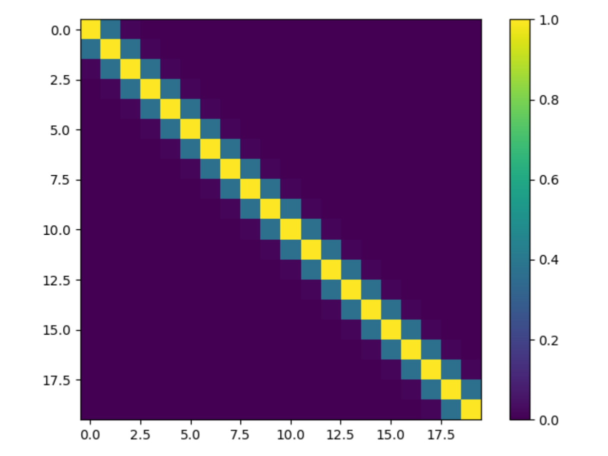
$$l = 2.0, \sigma = 1.0, x = [-3, \ldots, 3],$$

Kernel Parameters



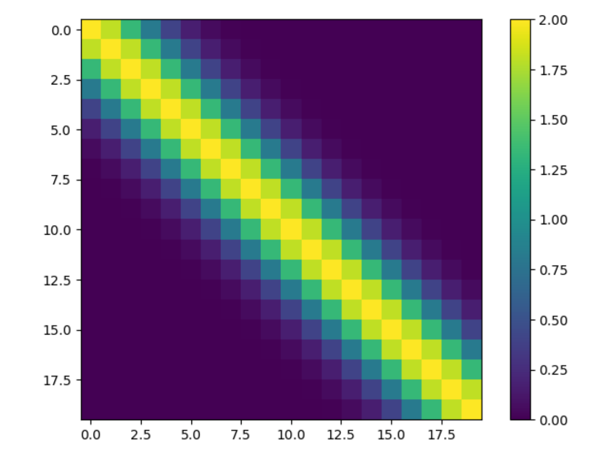
$$l = 4.0, \sigma = 1.0, x = [-3, \ldots, 3],$$

Kernel Parameters



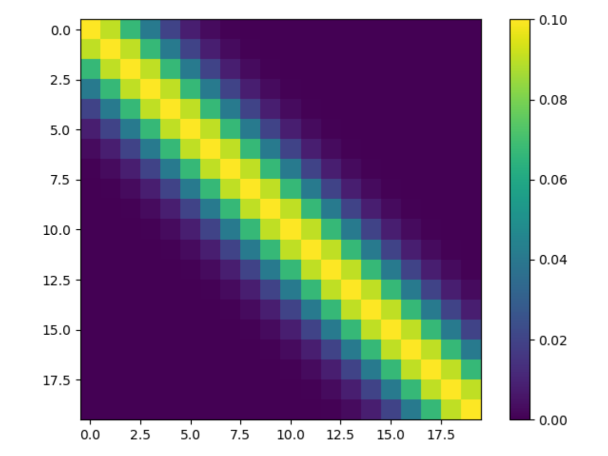
$$l = 0.1, \sigma = 1.0, x = [-3, \ldots, 3],$$

Kernel Parameters



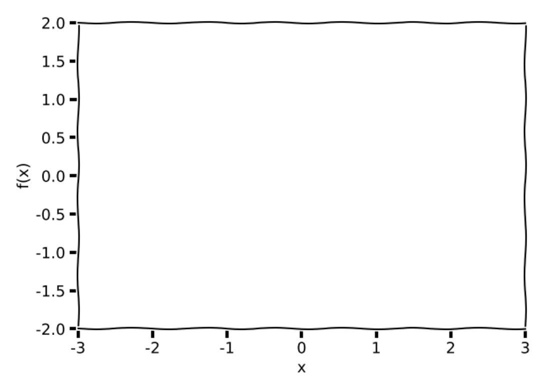
$$l = 1.0, \sigma = 2.0, x = [-3, \ldots, 3],$$

Kernel Parameters



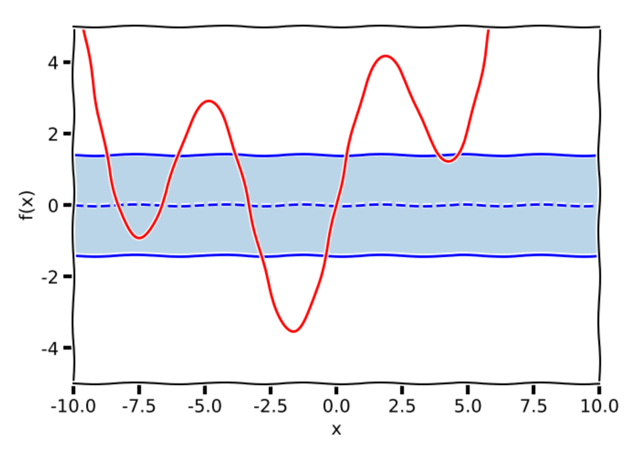
$$l = 1.0, \sigma = 0.1, x = [-3, \ldots, 3],$$

Why is this a prior?



$$p(\mathbf{f}|\mathbf{X}, \theta)$$

Marginalisation



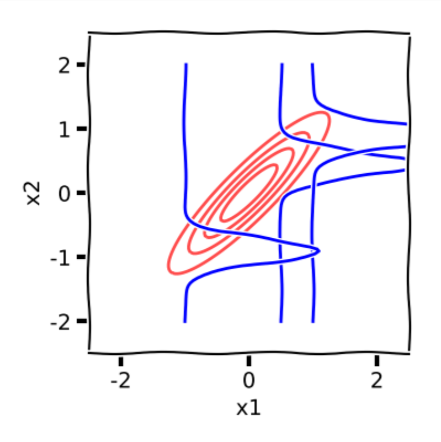
$$p(\mathbf{Y}|\mathbf{X}, \theta) = \int p(\mathbf{Y}|\mathbf{f})p(\mathbf{f}|\mathbf{X}, \theta)\mathrm{d}\mathbf{f}$$

Prediction

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} k(\mathbf{X},\mathbf{X}) & k(\mathbf{X},\mathbf{x}_*) \\ k(\mathbf{x}_*,\mathbf{X}) & k(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix} \right)$$

$$\begin{aligned} p(f_* | \mathbf{x}_*, \mathbf{X}, \mathbf{f}, \boldsymbol{\theta}) = \mathcal{N}(&k(\mathbf{x}_*,\mathbf{X})^{\mathrm{T}} K(\mathbf{X},\mathbf{X})^{-1}\mathbf{f}, \\ &k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,\mathbf{X})^{\mathrm{T}} K(\mathbf{X},\mathbf{X})^{-1} K(\mathbf{X},\mathbf{x}_*)) \end{aligned}$$

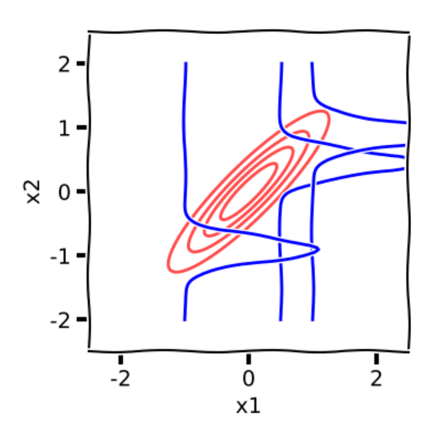
Conditional Gaussians



$$\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}\right)$$

$$p(y_1, y_2) = p(y_1|y_2)p(y_2)$$

Conditional Gaussians



$$\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}\right)$$

$$p(y_1|y_2) = \mathcal{N}(\mu_1 + \sigma_{21}\sigma_{22}^{-1}(y_2 - \mu_2), \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21})$$

Prediction

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} k(\mathbf{X},\mathbf{X}) & k(\mathbf{X},\mathbf{x}_*) \\ k(\mathbf{x}_*,\mathbf{X}) & k(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix} \right)$$

$$\begin{aligned} p(f_*|\mathbf{x}_*,\mathbf{X},\mathbf{f},\boldsymbol{\theta}) = \mathcal{N}(& k(\mathbf{x}_*,\mathbf{X})^{\mathrm{T}} K(\mathbf{X},\mathbf{X})^{-1}\mathbf{f}, \\ & k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,\mathbf{X})^{\mathrm{T}} K(\mathbf{X},\mathbf{X})^{-1} K(\mathbf{X},\mathbf{x}_*)) \end{aligned}$$

Intractability

$$p(\mathbf{Y}) = \int p(\mathbf{Y}|\mathbf{f})p(\mathbf{f}|\mathbf{X},\theta)p(\mathbf{X})p(\theta)\mathrm{d}\mathbf{f}\mathrm{d}\mathbf{X}\mathrm{d}\theta$$

$$p(\mathbf{X}) = \mathcal{N}(\mathbf{0},\mathbf{I}), \qquad p(f|\mathbf{X},\theta) = \mathcal{N}(\mathbf{0},k(\mathbf{X},\mathbf{X}))$$

$$\mathcal{N}(\mathbf{f}|\boldsymbol{\mu},K(\mathbf{X},\mathbf{X})) = \frac{1}{(2\pi)^{\frac{D}{2}}|K(\mathbf{X},\mathbf{X})|^{\frac{1}{2}}|} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}K(\mathbf{X},\mathbf{X})^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

$$K(\mathbf{x}_j,\mathbf{x}_i) = \sigma^2 e^{-(x_i-x_j)^{\mathrm{T}}(x_i-x_j)}$$

- if we can reach the model evidence we can write down all the probabilities that we want
- the above is analytically intractable for a non-linear covariance

Summary

Summary

eof

References