# 10-424/624: Bayesian Methods in ML Lecture 7: Supplement & Figures

Henry Chai

2/4/25

# Efficiency

- Depending on the transformation $\phi$ and the dimensionality of the original input space $\mathcal{X}$, computing $\phi(\boldsymbol{x})$ can be prohibitively computationally expensive
  - Computing $\phi_2(\boldsymbol{x}) = [x_1, x_2, \ldots, x_d, x_1^2, x_1 x_2, \ldots, x_d^2]$ requires $d + \binom{d}{2} + d = \frac{d^2+3d}{2} = O(d^2)$ time
  - Computing $\phi_{10}(\boldsymbol{x})$ requires $O(d^{10})$ time
- Tradeoff:
  - High-dimensional transformations can result in good hypotheses (as long as they don't overfit)
  - High-dimensional transformations are expensive

# The Kernel Trick

- Approach: instead of computing $\phi(\boldsymbol{x})$, find some function $k_\phi$ s.t. $k_\phi(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}')\ \forall\ \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$
  - $k_\phi(\boldsymbol{x}, \boldsymbol{x}')$ should be cheaper to compute than $\phi(\boldsymbol{x})$

- Example: $\tilde{\phi}_2(\boldsymbol{x}) = \left[x_1, \ldots, x_d, x_1^2, \sqrt{2}x_1x_2, \ldots, \sqrt{2}x_{d-1}x_d, x_d^2\right]$

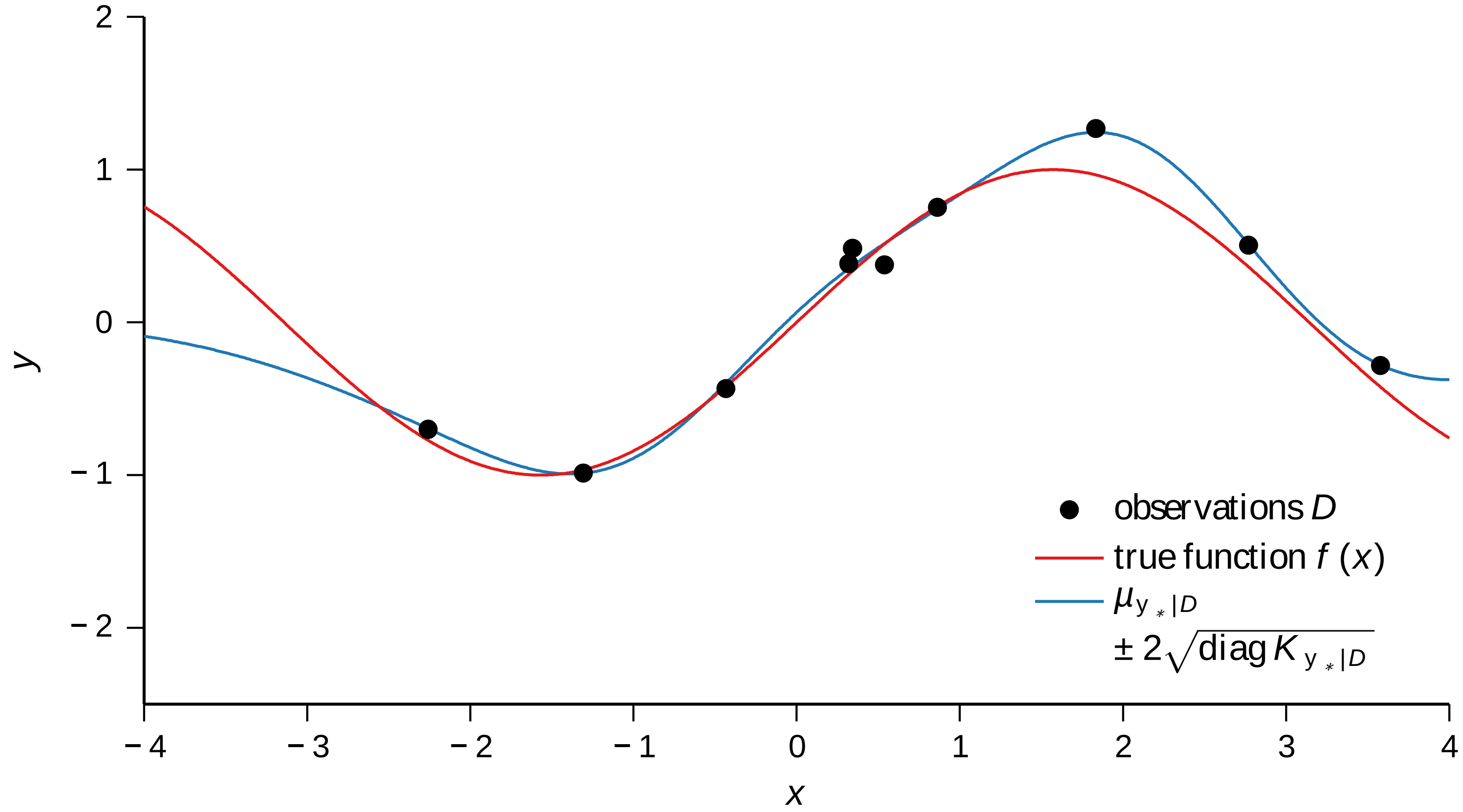
$$\tilde{\phi}_2(\boldsymbol{x})^T \tilde{\phi}_2(\boldsymbol{x}') = \sum_{i=1}^{d} x_i x_i' + \sum_{i=1}^{d} x_i^2 {x_i'}^2 + \sum_{i=1}^{d} \sum_{j \neq i} 2 x_i x_i' x_j x_j'$$

$$= \sum_{i=1}^{d} x_i x_i' + \left(\sum_{i=1}^{d} x_i x_i'\right)^2 = \boldsymbol{x}^T \boldsymbol{x}' + (\boldsymbol{x}^T \boldsymbol{x}')^2$$

$$k_{\tilde{\phi}_2}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T \boldsymbol{x}' + (\boldsymbol{x}^T \boldsymbol{x}')^2$$

- Computing $\tilde{\phi}_2(\boldsymbol{x})^T \tilde{\phi}_2(\boldsymbol{x}')$ requires $O(d^2)$ time whereas computing $k_{\tilde{\phi}_2}(\boldsymbol{x}, \boldsymbol{x}')$ only takes $O(d)$!

# Squared Exponential Kernel




Bayesian linear regression using the squared exponential kernel $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \lambda^2 \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2}{2\ell^2}\right)$ with $\lambda = \ell = 1$.

The true function is $f(x) = \sin x$ with additive, zero-mean Gaussian noise where the noise variance is $\sigma^2 = 0.1^2$