UNIVERSITY OF AMSTERDAM
Informatics Institute

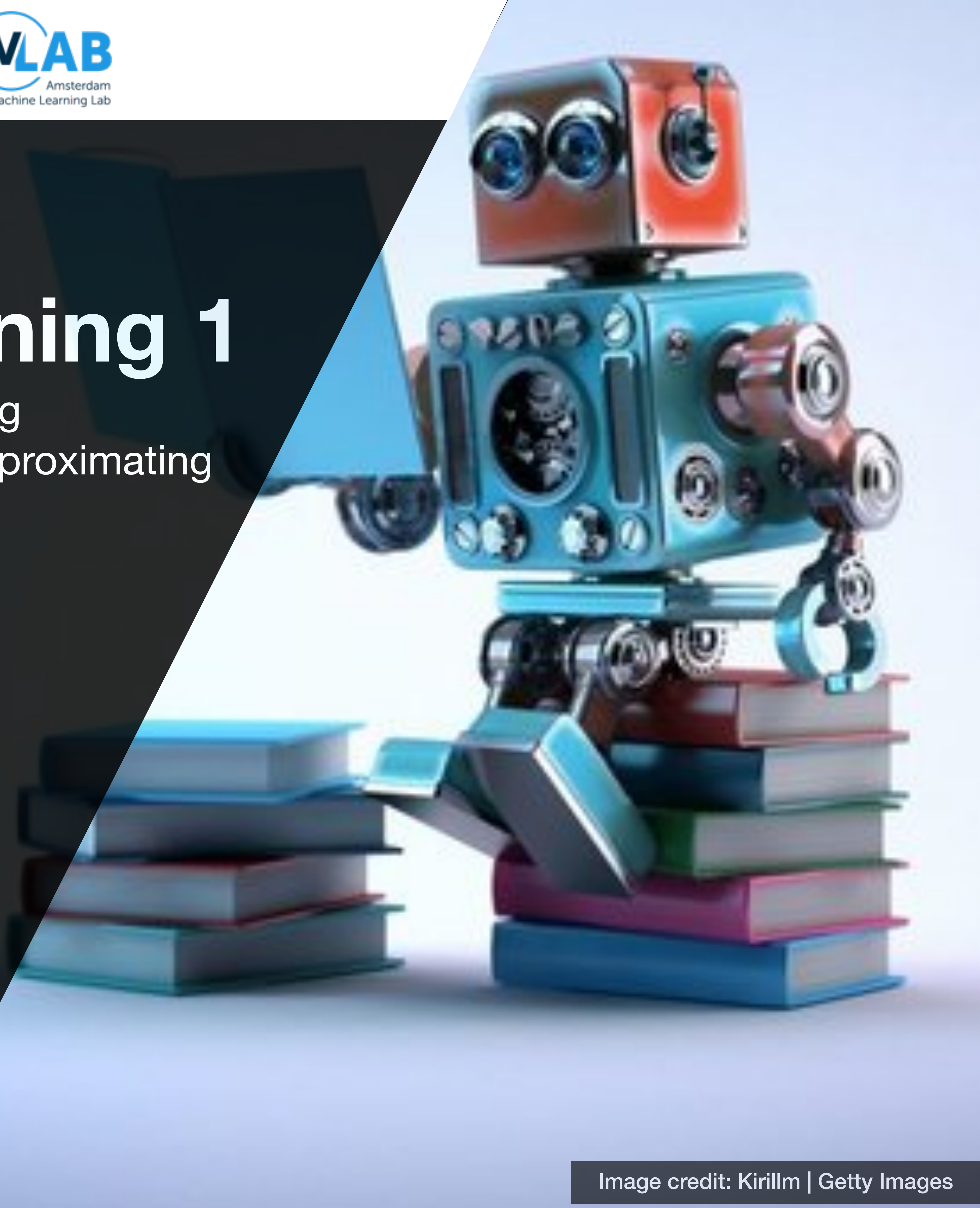
AMLAB
Amsterdam Machine Learning Lab

Machine Learning 1

Lecture 5.3 - Supervised Learning
Bayesian Linear Regression - Approximating the Model Evidence

Erik Bekkers

(Bishop 3.5.0)



Slide credits: Patrick Forré and Rianne van den Berg

Image credit: Kirillm | Getty Images

Model Evidence for Linear Basis Models

- Full Bayesian treatment to model evidence:

$$p(\mathbf{t}|\mathbf{X},\mathcal{M}_i) = \iiint \underline{p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta,\mathcal{M}_i)p(\mathbf{w}|\mathbf{t},\mathbf{X},\alpha,\mathcal{M}_i)}p(\alpha,\beta|\mathbf{t},\mathbf{X},\mathcal{M}_i)\mathrm{d}\mathbf{w}\mathrm{d}\alpha\mathrm{d}\beta$$

$$= \iint p(\underline{t}|X,\alpha,\beta,\mathcal{M}_i)\underbrace{p(\alpha,\beta|\underline{t},X,\mathcal{M}_i)}_{\text{assume sharply peaked at } \alpha^*,\beta^*}d\alpha\, d\beta$$

- Empirical Bayes / evidence approximation:

done iterative way Bishop 3.5.2

$$p(\mathbf{t}|\mathbf{X},\mathcal{M}_i) \approx p(\underline{t}|X,\alpha^*,\beta^*,\mathcal{M}_i)$$

$$\alpha^*,\beta^* = \underset{\alpha,\beta}{\operatorname{argmax}}\ p(\alpha,\beta|\underline{t},X,\mathcal{M}_i) \approx \underset{\alpha,\beta}{\operatorname{argmax}}\ p(\underline{t}|X,\alpha,\beta,\mathcal{M}_i)\cancel{p(\alpha,\beta,\mathcal{M}_i)}$$

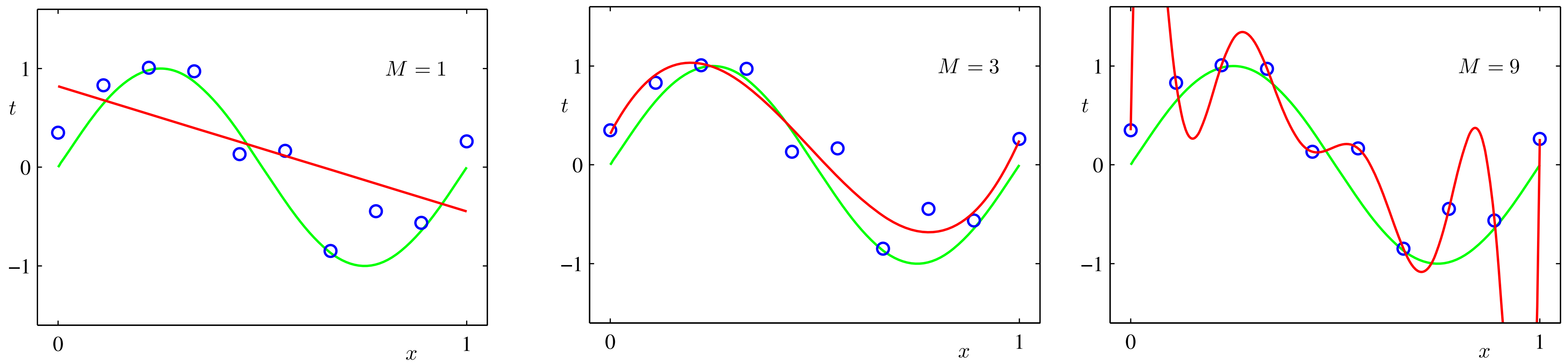
model selection

- Predictive distribution:

$$p(t'|\mathbf{x}',\mathbf{t},\mathbf{X},\mathcal{M}_i) \approx p(t'|\underline{x}',\underline{t},X,\alpha^*,\beta^*,\mathcal{M}_i)$$

Polynomial Regression: Choosing M

Model complexity: trade-off between model fit and model complexity



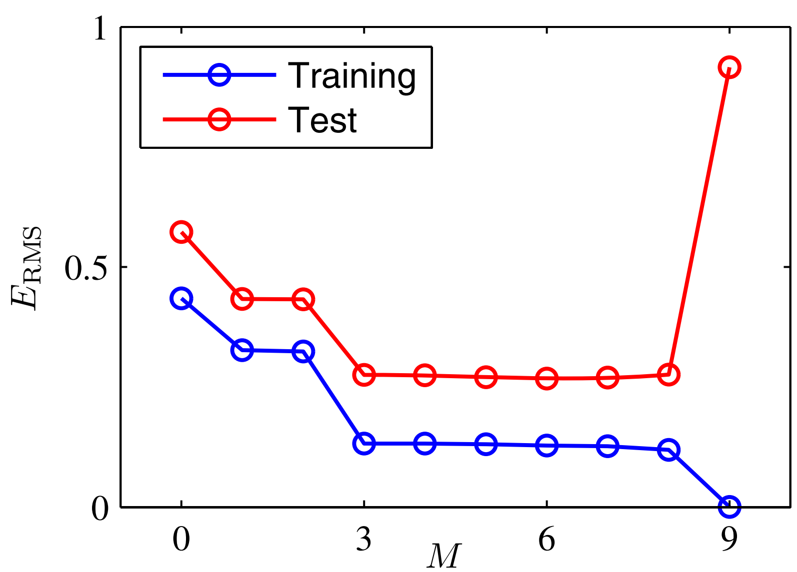


Figure: E_{rmse} (Bishop 1.5)

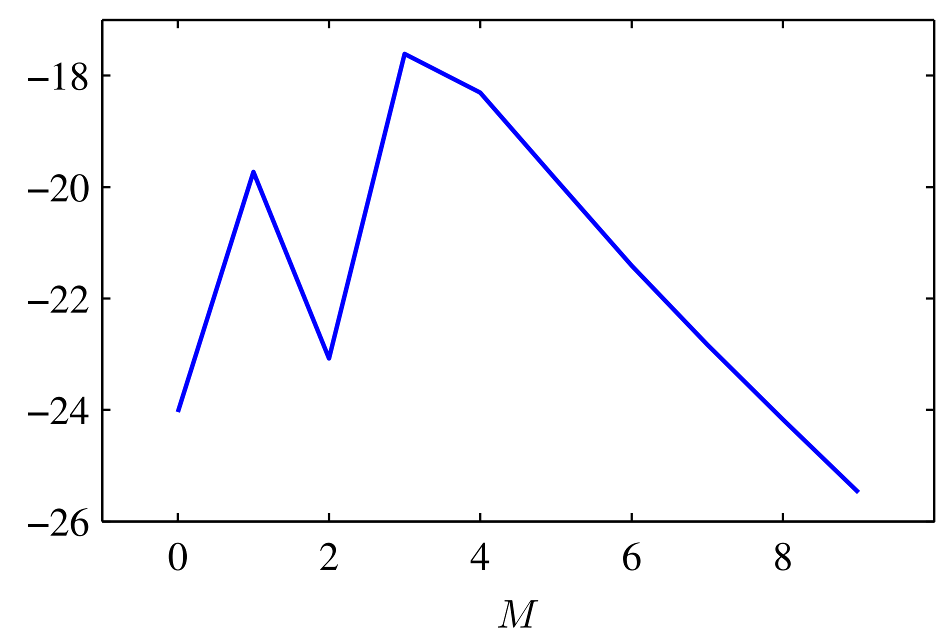


Figure: log model evidence (Bishop 3.14)