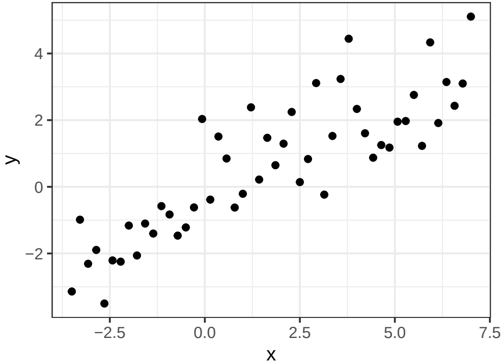
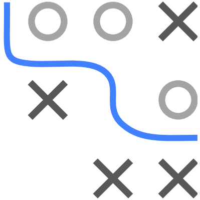
REVIEW: THE BAYESIAN LINEAR MODEL

Let $\mathcal{D} = \left\{(\mathbf{x}^{(1)}, y^{(1)}), ..., (\mathbf{x}^{(n)}, y^{(n)})\right\}$ be a training set of i.i.d. observations from some unknown distribution.

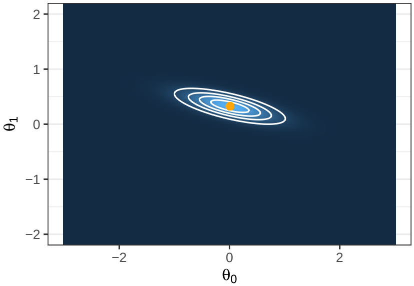


Let $\mathbf{y} = (y^{(1)}, ..., y^{(n)})^\top$ and $\mathbf{X} \in \mathbb{R}^{n \times p+1}$ be the design matrix where the i-th row contains vector $\mathbf{x}^{(i)}$.

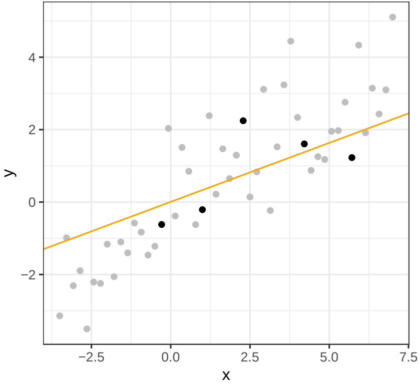


REVIEW: THE BAYESIAN LINEAR MODEL

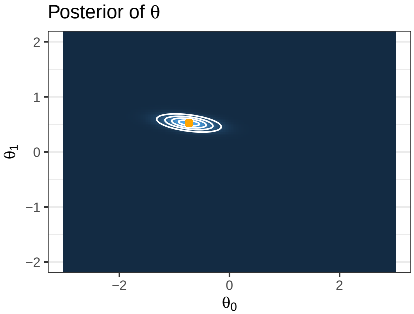
Posterior of θ

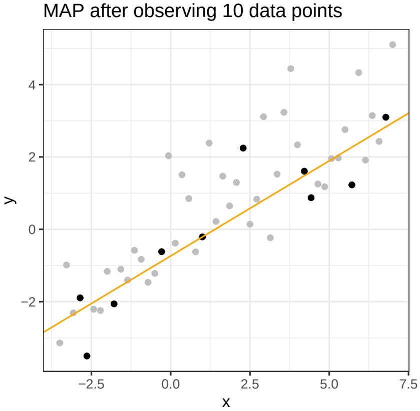


MAP after observing 5 data points

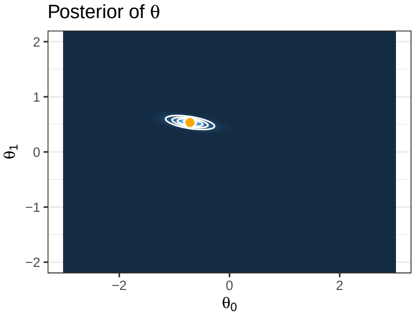


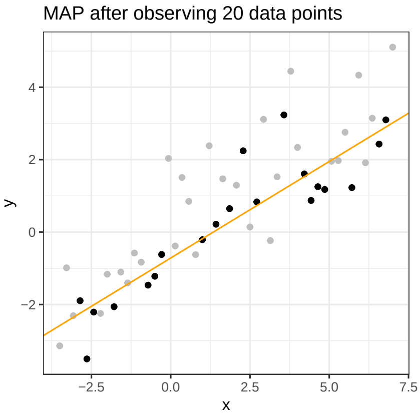
REVIEW: THE BAYESIAN LINEAR MODEL





REVIEW: THE BAYESIAN LINEAR MODEL





REVIEW: THE BAYESIAN LINEAR MODEL

Based on the posterior distribution

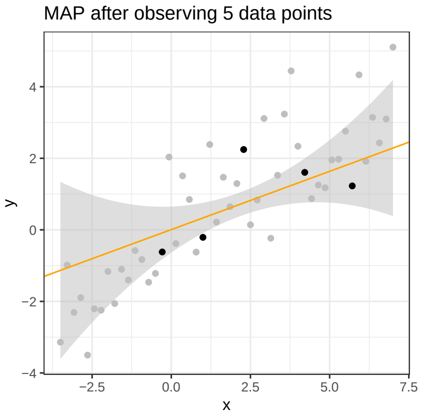
$$\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{y} \sim \mathcal{N}(\sigma^{-2}\boldsymbol{A}^{-1}\mathbf{X}^\top \mathbf{y}, \boldsymbol{A}^{-1})$$

we can derive the predictive distribution for a new observation $\mathbf{x}_*$. The predictive distribution for the Bayesian linear model, i.e. the distribution of $\boldsymbol{\theta}^\top \mathbf{x}_*$, is

$$y_* \mid \mathbf{X}, \mathbf{y}, \mathbf{x}_* \sim \mathcal{N}(\sigma^{-2}\mathbf{y}^\top \mathbf{X}\mathbf{A}^{-1}\mathbf{x}_*, \mathbf{x}_*^\top \mathbf{A}^{-1}\mathbf{x}_*)$$

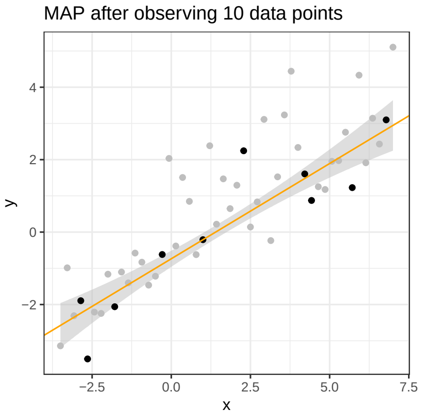
Please see the Deep Dive part for more details.

REVIEW: THE BAYESIAN LINEAR MODEL



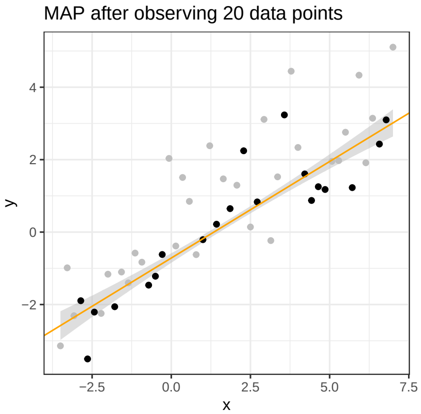
For every test input $\mathbf{x}_*$, we get a distribution over the prediction y_*. In particular, we get a posterior mean (orange) and a posterior variance (grey region equals $+/-$ two times standard deviation).

REVIEW: THE BAYESIAN LINEAR MODEL



For every test input $\mathbf{x}_*$, we get a distribution over the prediction y_*. In particular, we get a posterior mean (orange) and a posterior variance (grey region equals $+/-$ two times standard deviation).

REVIEW: THE BAYESIAN LINEAR MODEL



For every test input $\mathbf{x}_*$, we get a distribution over the prediction y_*. In particular, we get a posterior mean (orange) and a posterior variance (grey region equals $+/-$ two times standard deviation).

SUMMARY: THE BAYESIAN LINEAR MODEL

- By switching to a Bayesian perspective, we do not only have point estimates for the parameter $\boldsymbol{\theta}$, but whole **distributions**
- From the posterior distribution of $\boldsymbol{\theta}$, we can derive a predictive distribution for $y_* = \boldsymbol{\theta}^\top \mathbf{x}_*$.
- We can perform online updates: Whenever datapoints are observed, we can update the **posterior distribution** of $\boldsymbol{\theta}$

Next, we want to develop a theory for general shape functions, and not only for linear function.