Introduction to Bayesian Machine Learning

Rajbir-Singh Nirwan

July 30, 2019

Outline

- Why Probabilistic Modelling?
- Linear Regression
- Bayesian Linear Regression
- Classification
- Summary

Probabilistic Modelling

- Handling uncertainty by using probabilities
- Averaging over different possibilities

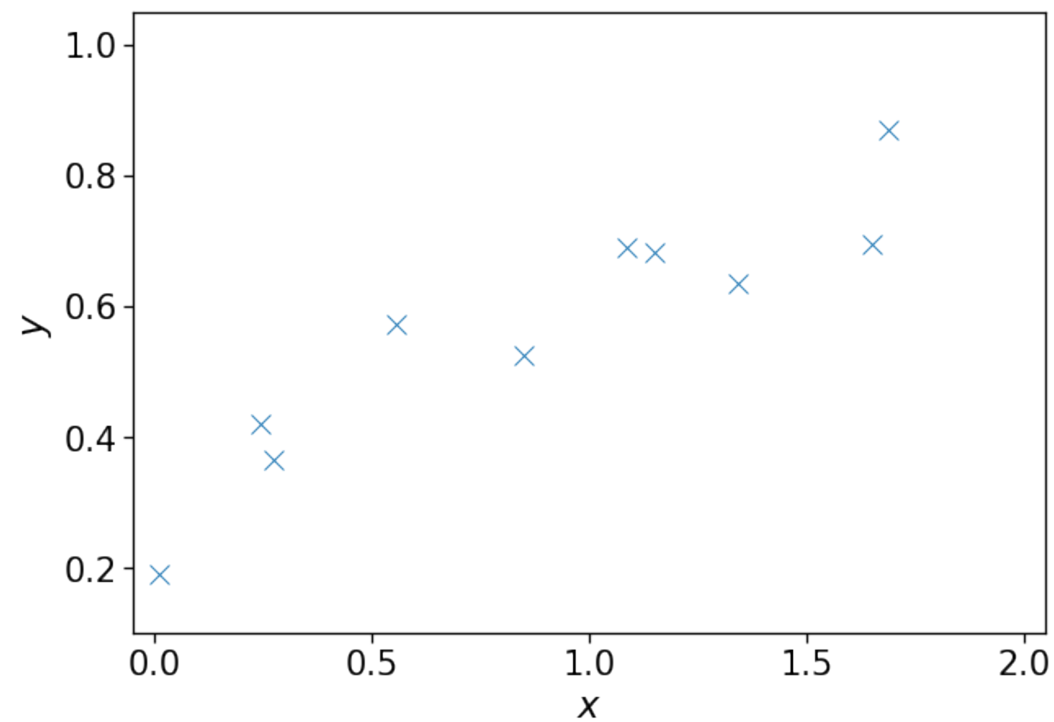
$$p(x) = \sum_y p(x, y) \qquad \& \qquad p(x, y) = p(y \mid x)\, p(x)$$

Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

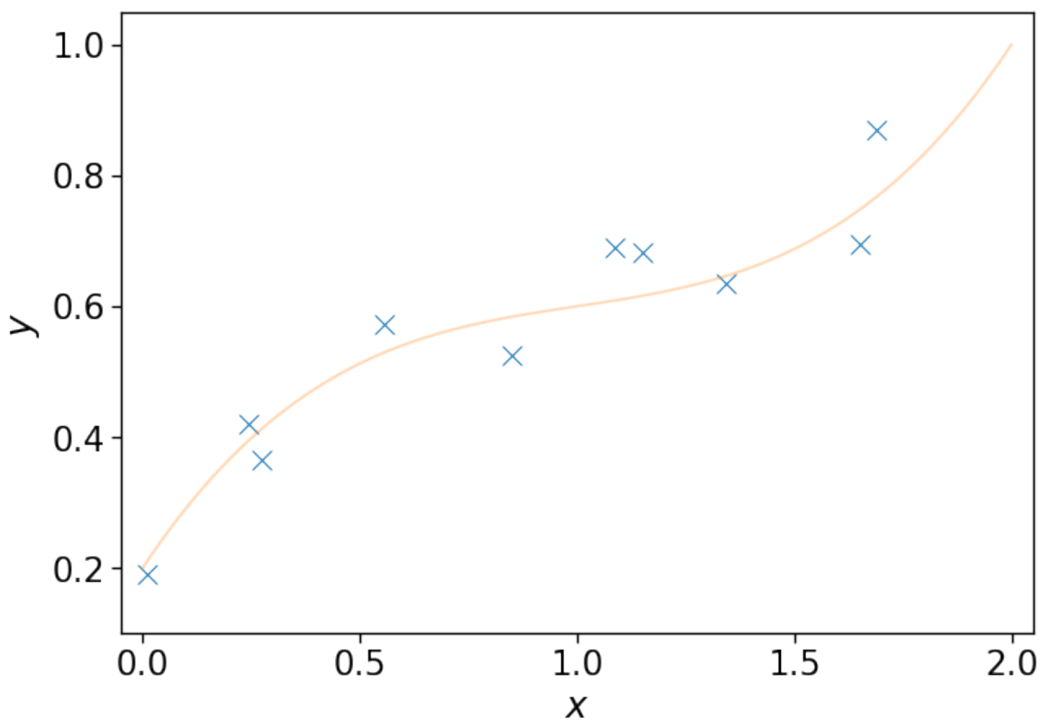


Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

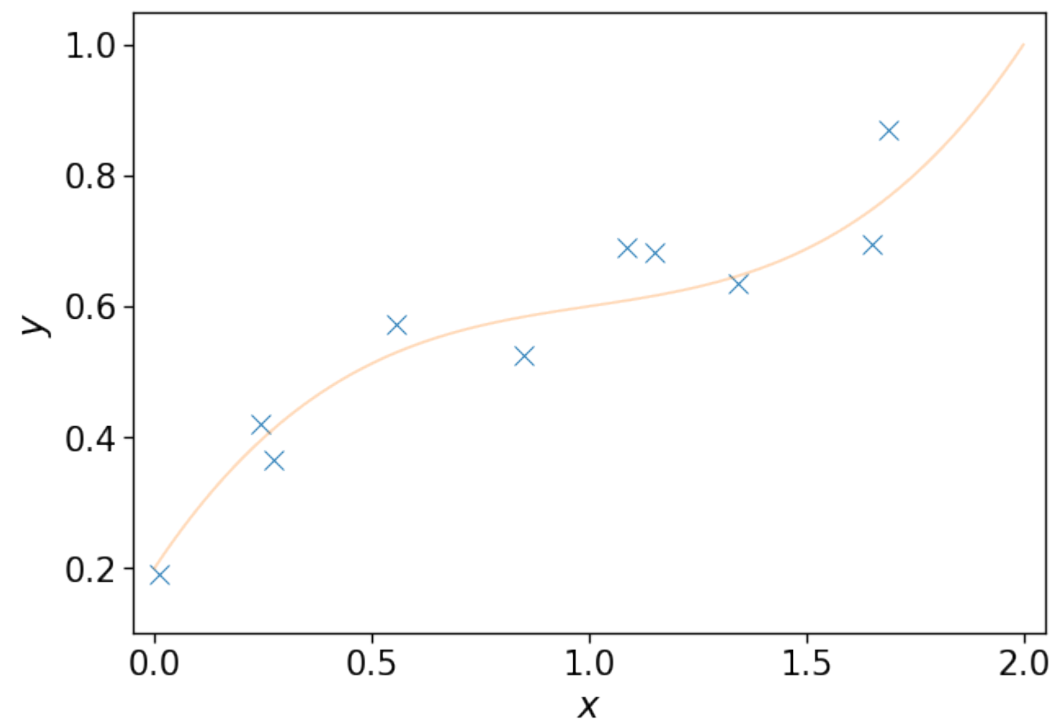
$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



Linear Regression

Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$



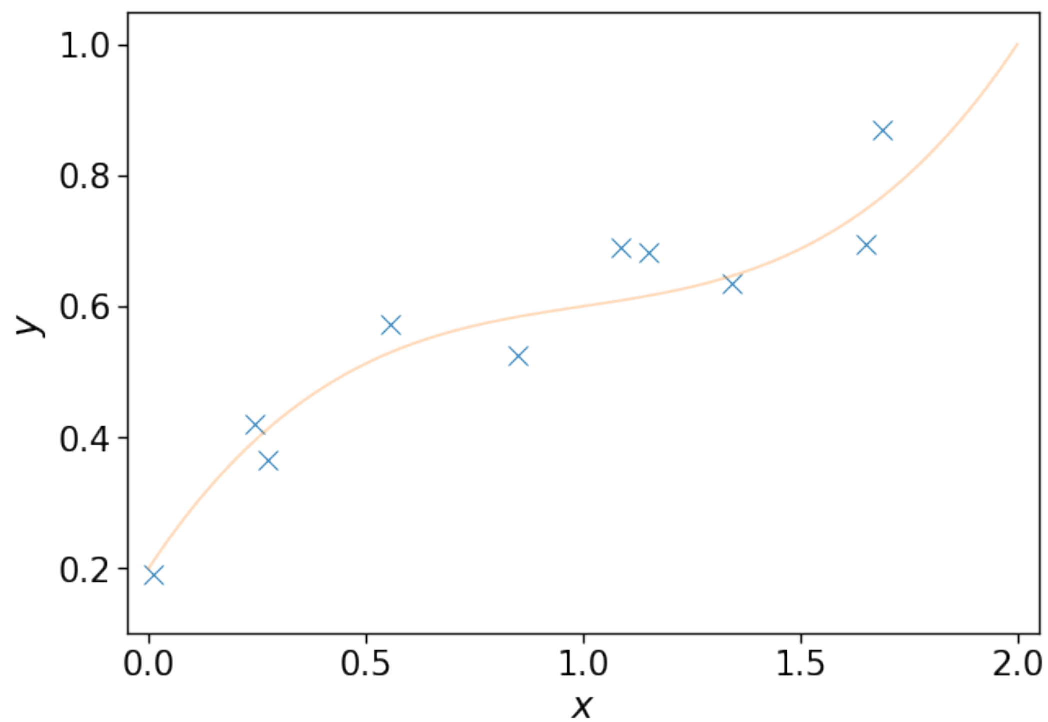
Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$

Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$f_{\boldsymbol{w}}(x) = \boldsymbol{w}_0 + \boldsymbol{w}_1 x$$

$$f_{\boldsymbol{w}}(x) = \boldsymbol{w}_0 + \boldsymbol{w}_1 x + \boldsymbol{w}_2 x^2$$

$$f_{\boldsymbol{w}}(x) = \boldsymbol{w}_0 + \boldsymbol{w}_1 x + \boldsymbol{w}_2 x^2 + \boldsymbol{w}_3 x^3$$

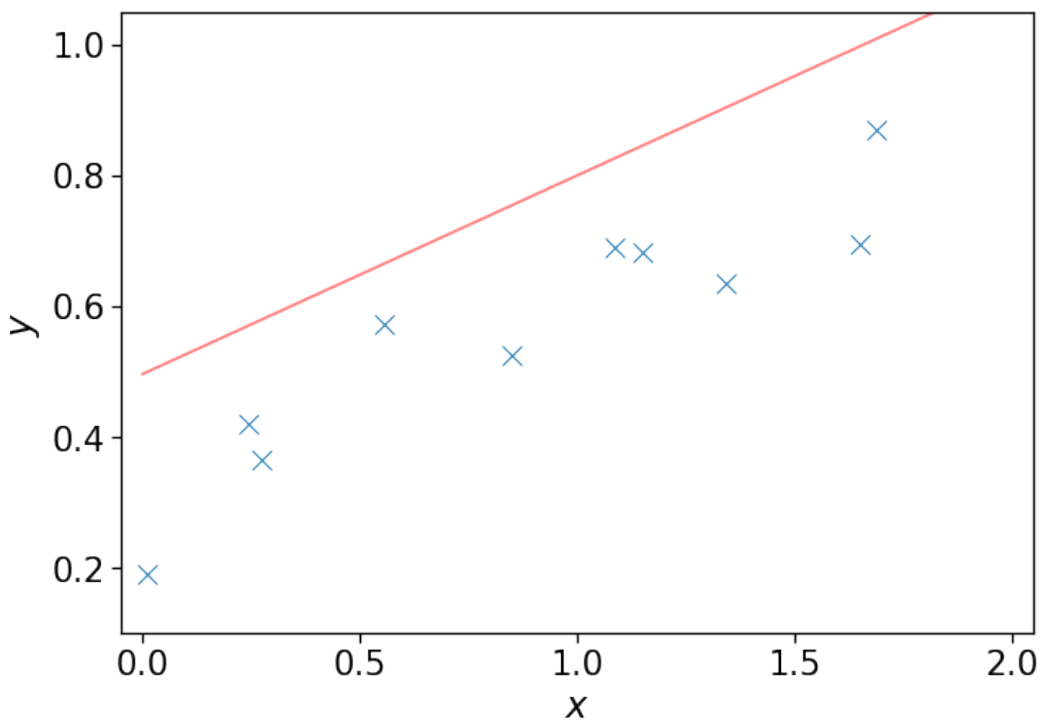
$$\rightarrow f_{\boldsymbol{w}}(\boldsymbol{\phi}) = \boldsymbol{w}^T \boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{w}_0 \\ \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \boldsymbol{w}_3 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \end{pmatrix}$$

Linear Regression

Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$

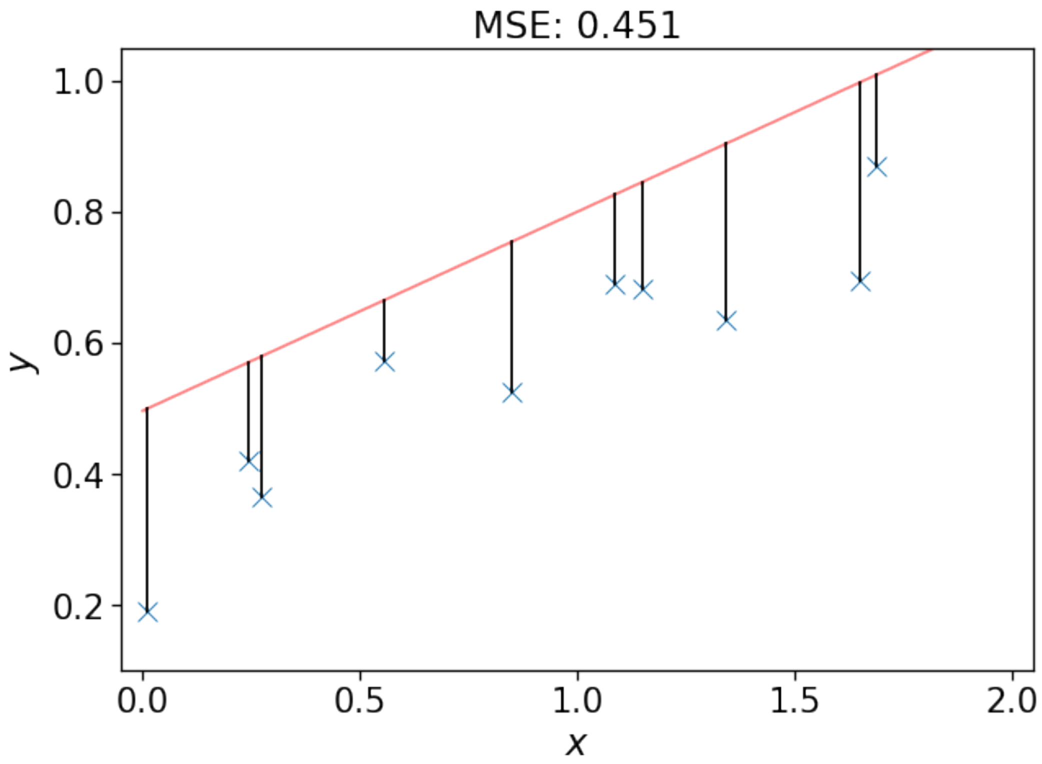
Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$



Linear Regression

Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$



Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$

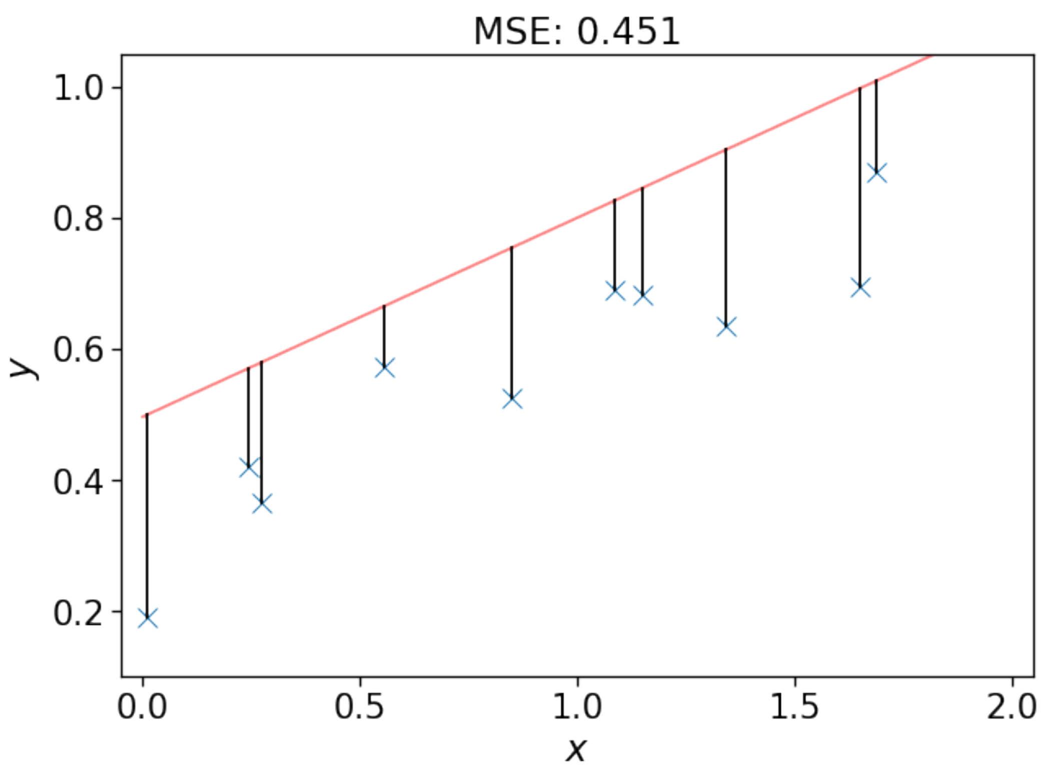
Linear Regression

Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$

Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$

$$E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N} \left(\boxed{y_n} - \boxed{\boldsymbol{w}^T \boldsymbol{x}_n}\right)^2$$



Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

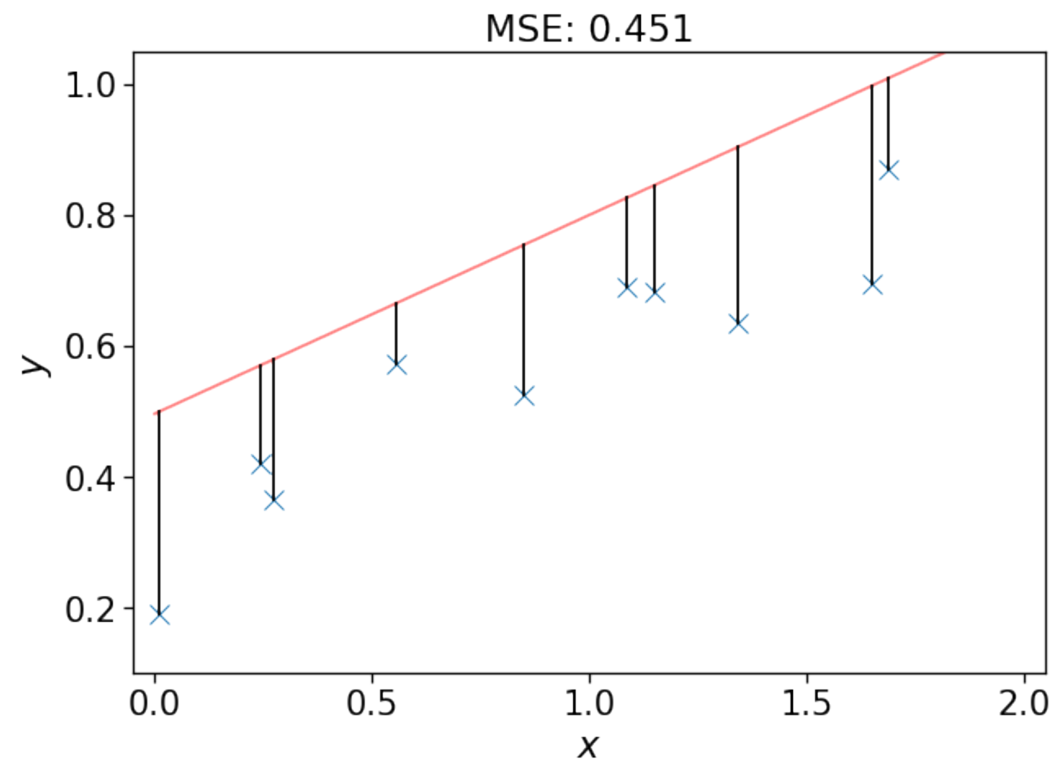
Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N}\left(\boxed{y_n} - \boxed{\boldsymbol{w}^T\boldsymbol{x}_n}\right)^2$$

Fit

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$$

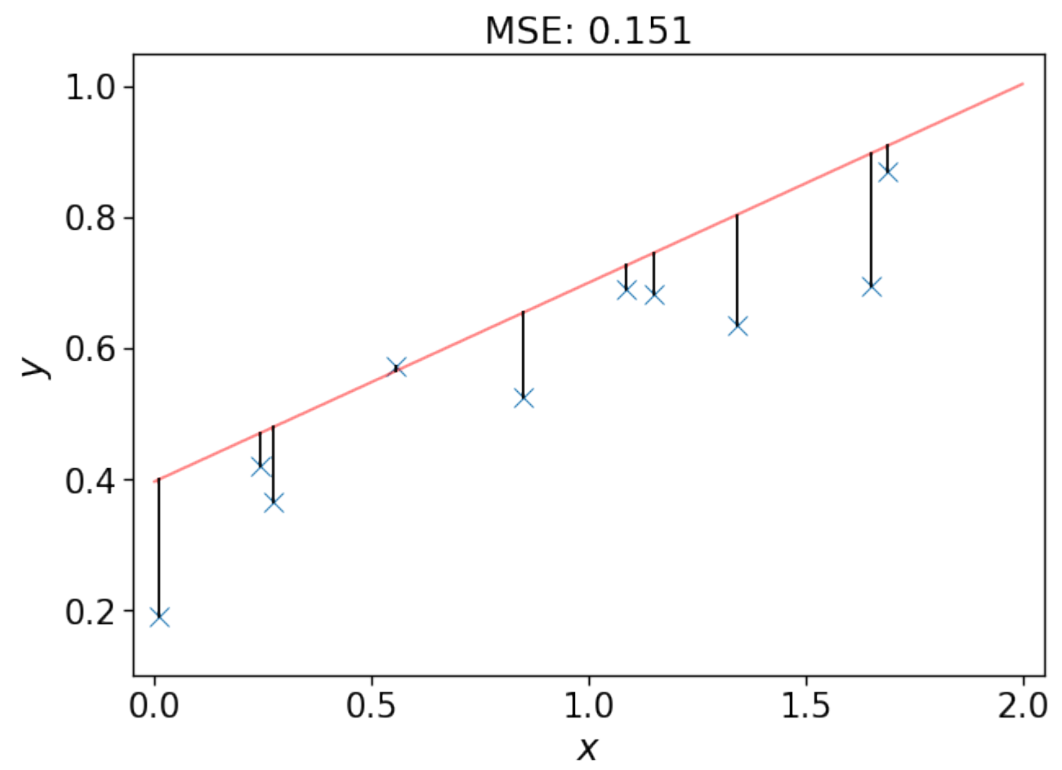


Linear Regression

Observed Data

$$\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$E_{\mathcal{D}}(\boldsymbol{w}) = \sum_{n=1}^{N} \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2$$

Fit

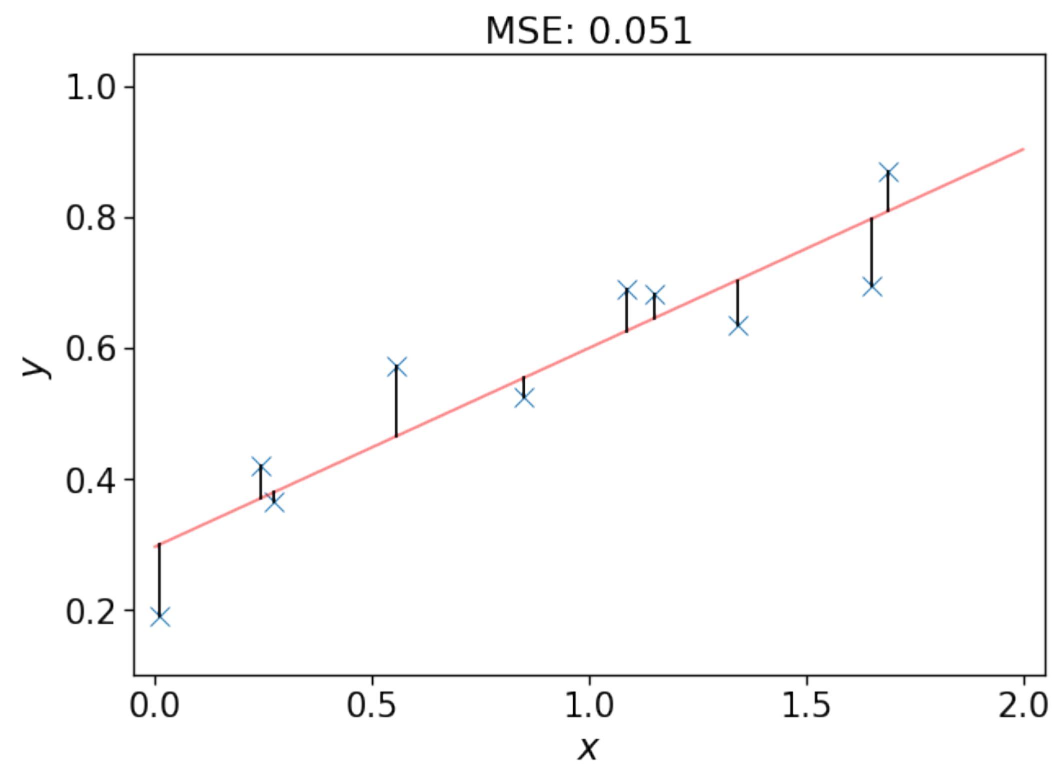
$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathcal{D}}(\boldsymbol{w})$$

Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N} \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2$$

Fit

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$$

Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

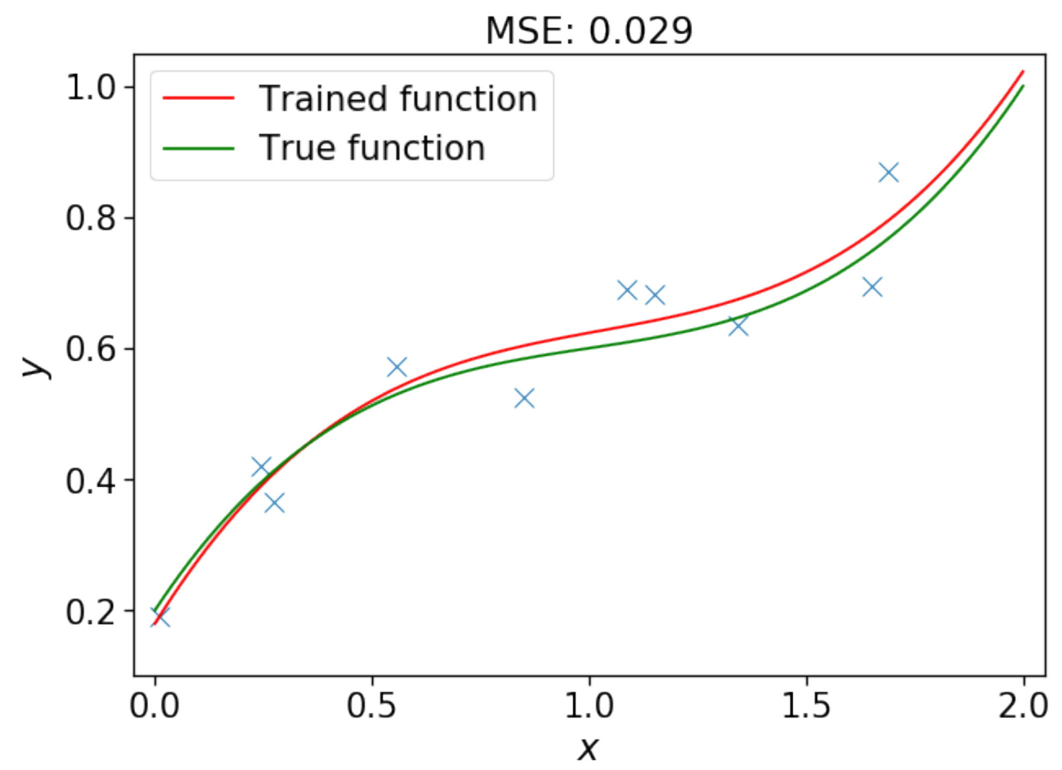
Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2$$

Fit

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$$



Linear Regression

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

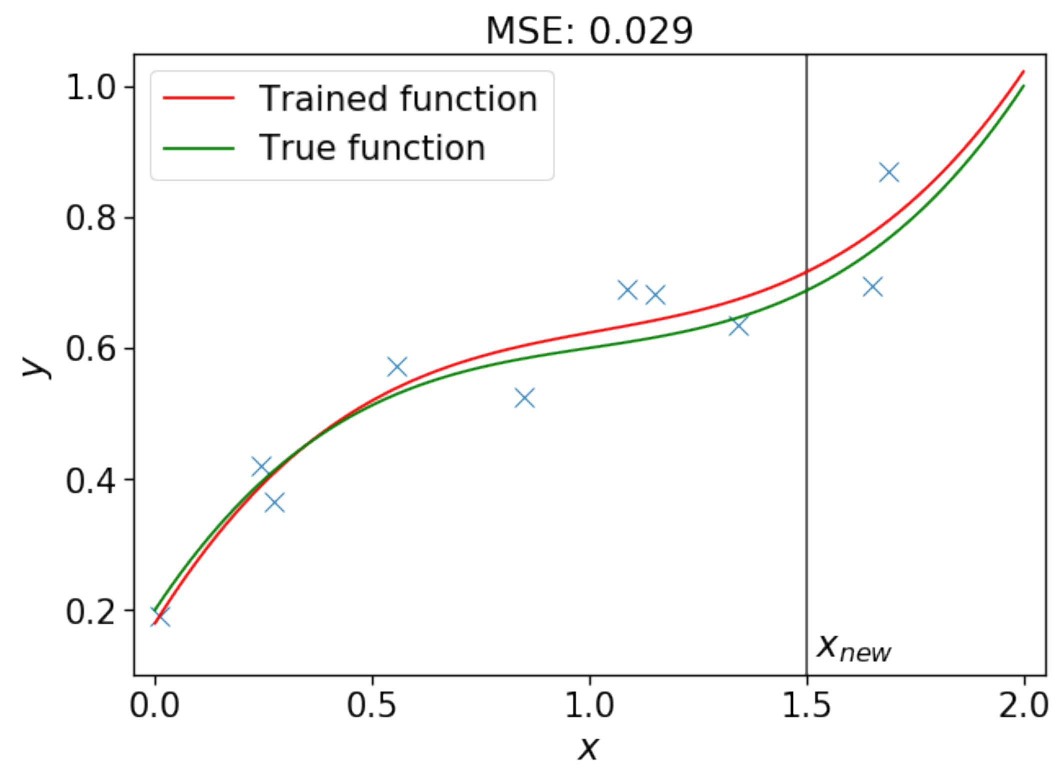
$$E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N} \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2$$

Fit

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$$

Prediction

$$y_{new} = f_{\boldsymbol{w}^*}(\boldsymbol{x}_{new})$$



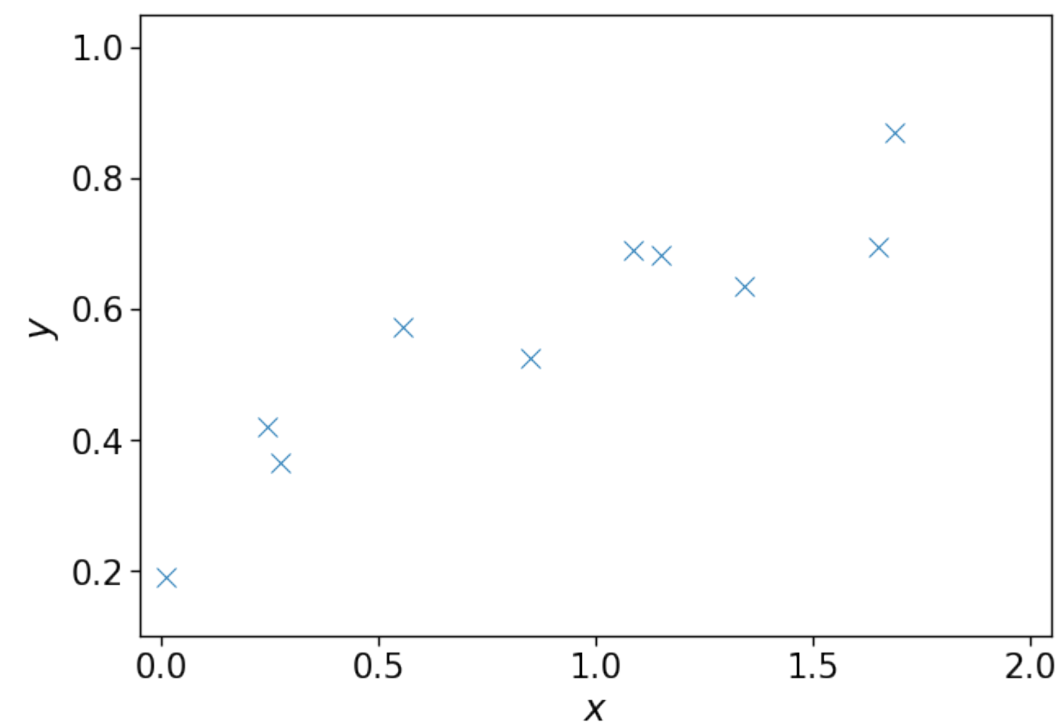
Bayesian Linear Regression

$$p(a) = \int p(a, b) \mathbf{d} b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



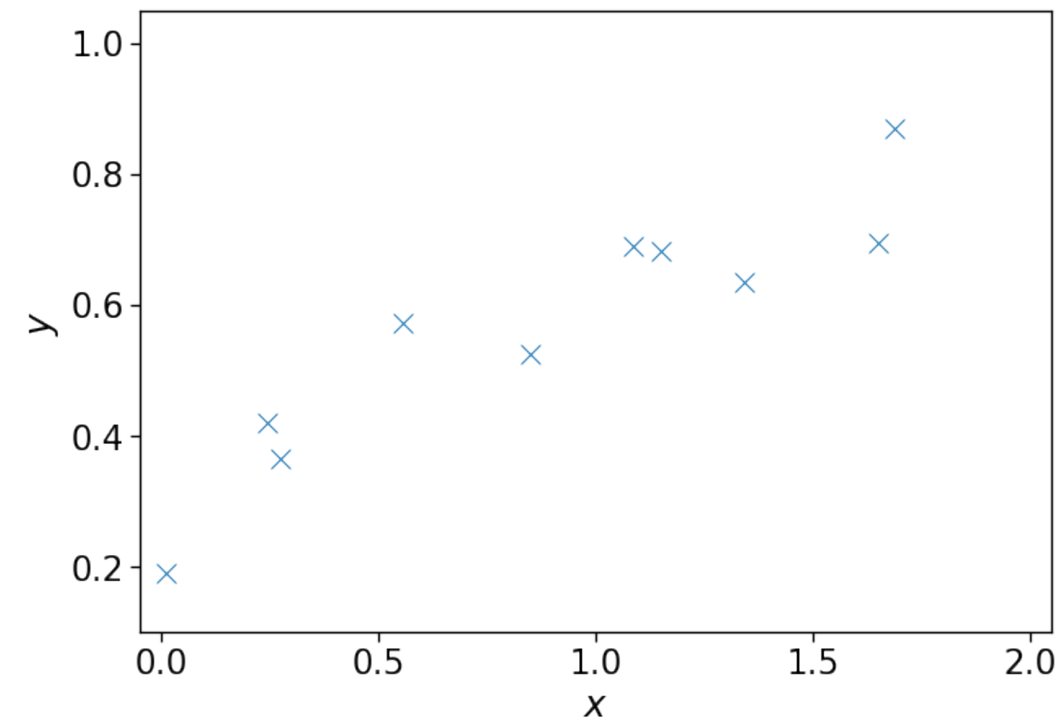
Bayesian Linear Regression

$$p(a) = \int p(a, b) \mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data $\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$

Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$



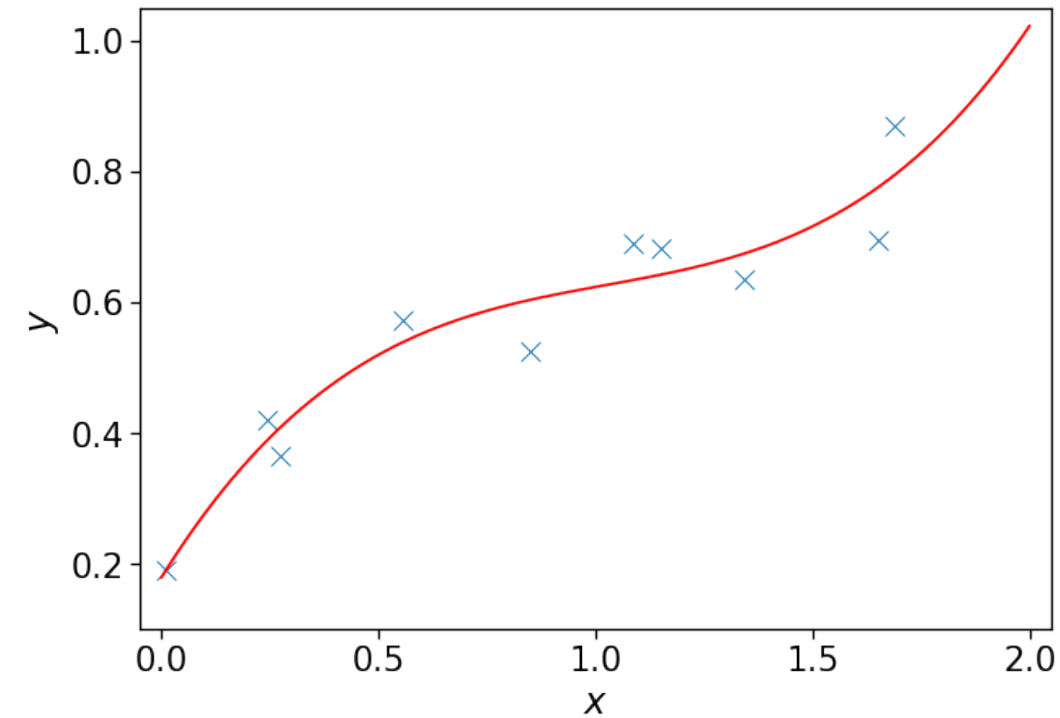
Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \,|\, b)\; p(b)$$

Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$

Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$



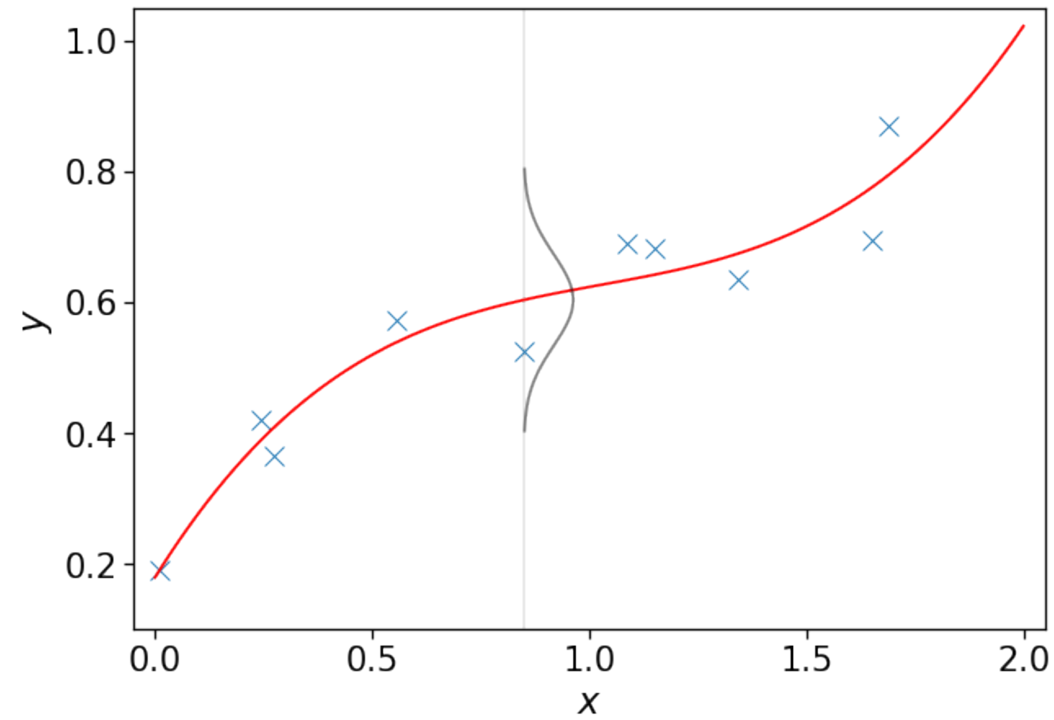
Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$

Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$



Bayesian Linear Regression

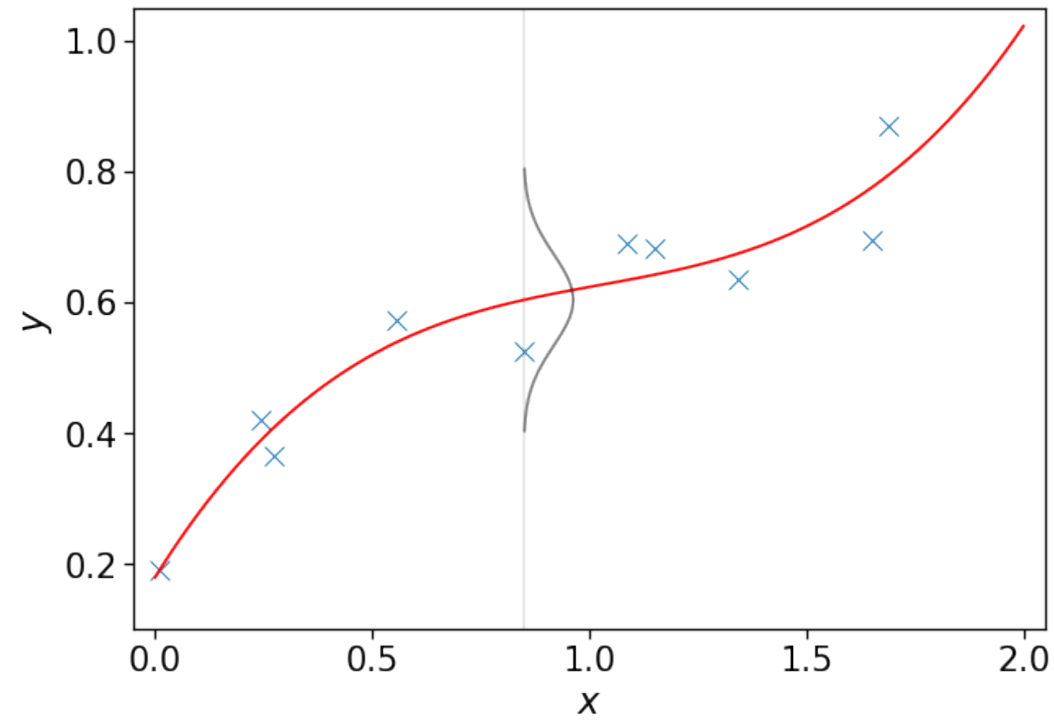
$$p(a) = \int p(a,b)\mathbf{d}b \qquad \& \qquad p(a,b) = p(a\,|\,b)\;p(b)$$

Observed Data $\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$

Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$

$$p(y_n\,|\,\boldsymbol{w},\sigma^2) = \mathcal{N}(y_n\,|\,\boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$



Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

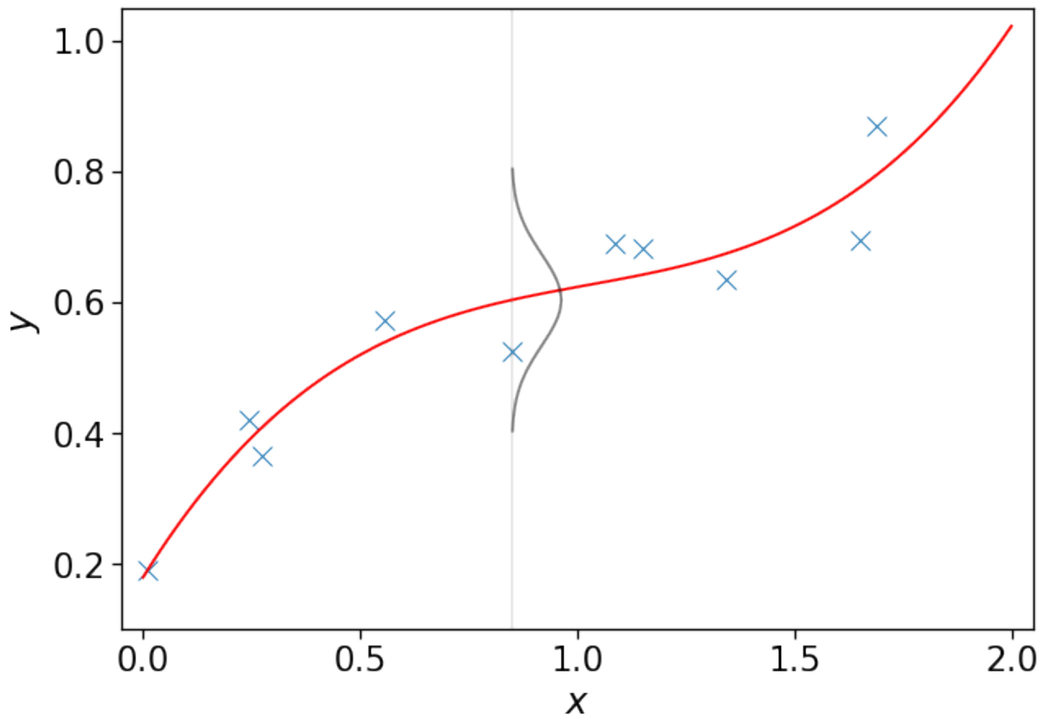
Observed Data $\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$

$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$

Model $y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathscr{N}(y_n \mid \boldsymbol{w}^T \boldsymbol{x}_n, \sigma^2)$$

$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^N p(y_n \mid \boldsymbol{w}, \sigma^2)$$



Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \,|\, b)\; p(b)$$

Observed Data

$$\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

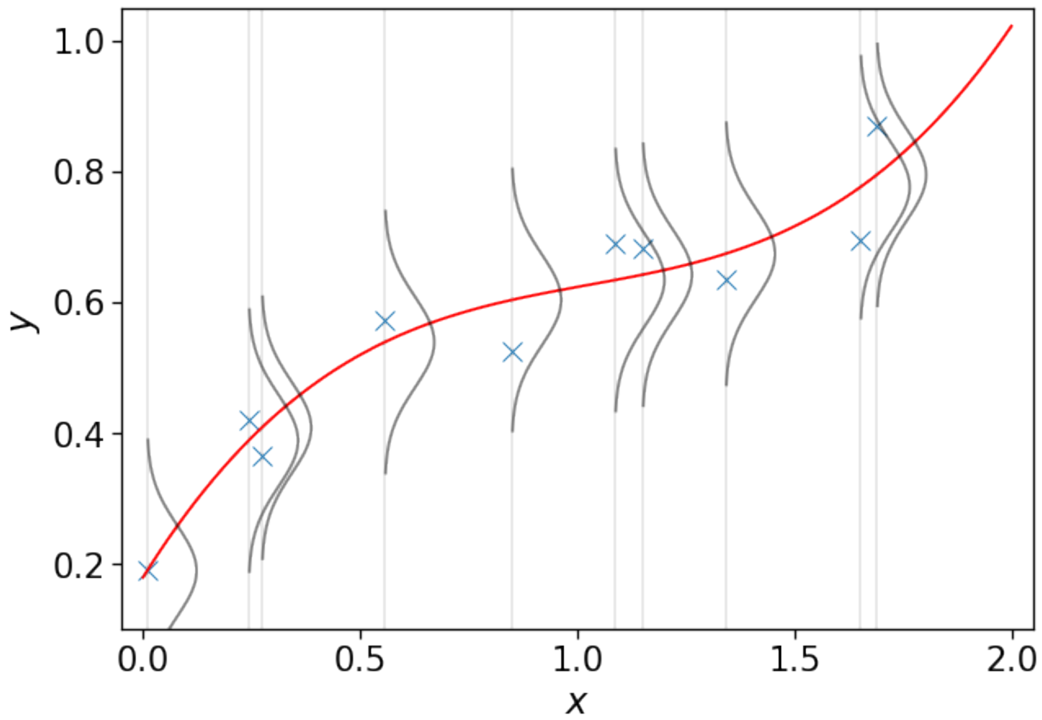
$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \,|\, \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \,|\, \boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

$$p(\boldsymbol{y} \,|\, \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \,|\, \boldsymbol{w}, \sigma^2)$$



Bayesian Linear Regression

$$p(a) = \int p(a,b)\mathbf{d}b \qquad \& \qquad p(a,b) = p(a\,|\,b)\;p(b)$$

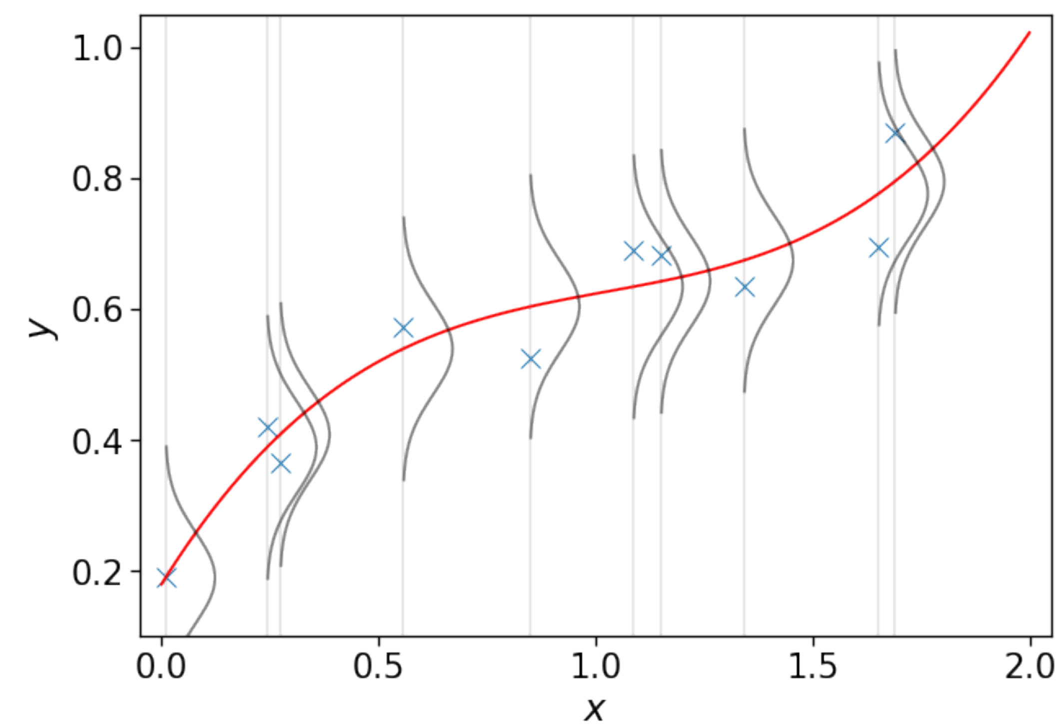
Observed Data $\quad \mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$

Model $\quad y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$

$$p(y_n\,|\,\boldsymbol{w},\sigma^2) = \mathcal{N}(y_n\,|\,\boldsymbol{w}^T\boldsymbol{x}_n,\sigma^2)$$

$$p(\boldsymbol{y}\,|\,\boldsymbol{w},\sigma^2) = \prod_{n=1}^{N} p(y_n\,|\,\boldsymbol{w},\sigma^2)$$



$$\log p\left(\boldsymbol{y}\,|\,\boldsymbol{w},\sigma^2\right) = \sum_{n=1}^{N}\left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2\right] = c_1 - c_2\sum_{n=1}^{N}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2$$

Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \,|\, b)\; p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

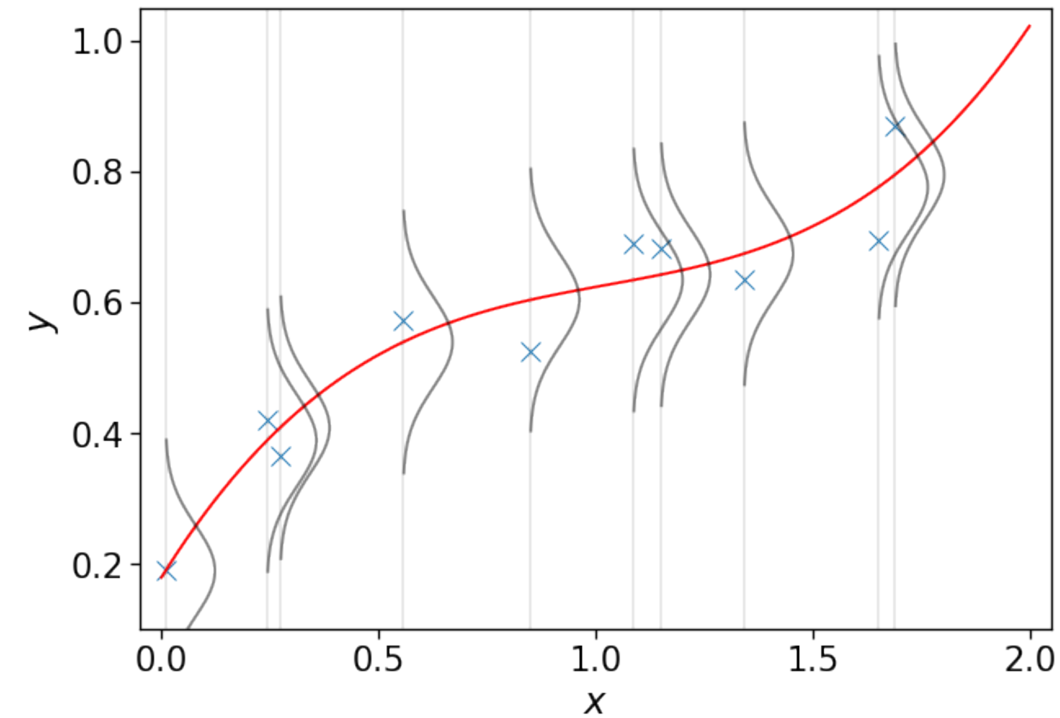
$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \,|\, \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \,|\, \boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

$$p(\boldsymbol{y} \,|\, \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \,|\, \boldsymbol{w}, \sigma^2)$$



$$\log p\left(\boldsymbol{y} \,|\, \boldsymbol{w}, \sigma^2\right) = \sum_{n=1}^{N}\left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2\right] = c_1 - c_2\sum_{n=1}^{N}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2$$

$$\max_{\boldsymbol{w}} \log p\left(\boldsymbol{y} \,|\, \boldsymbol{w}, \sigma^2\right) = \min_{\boldsymbol{w}} \sum_{n=1}^{N}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2$$

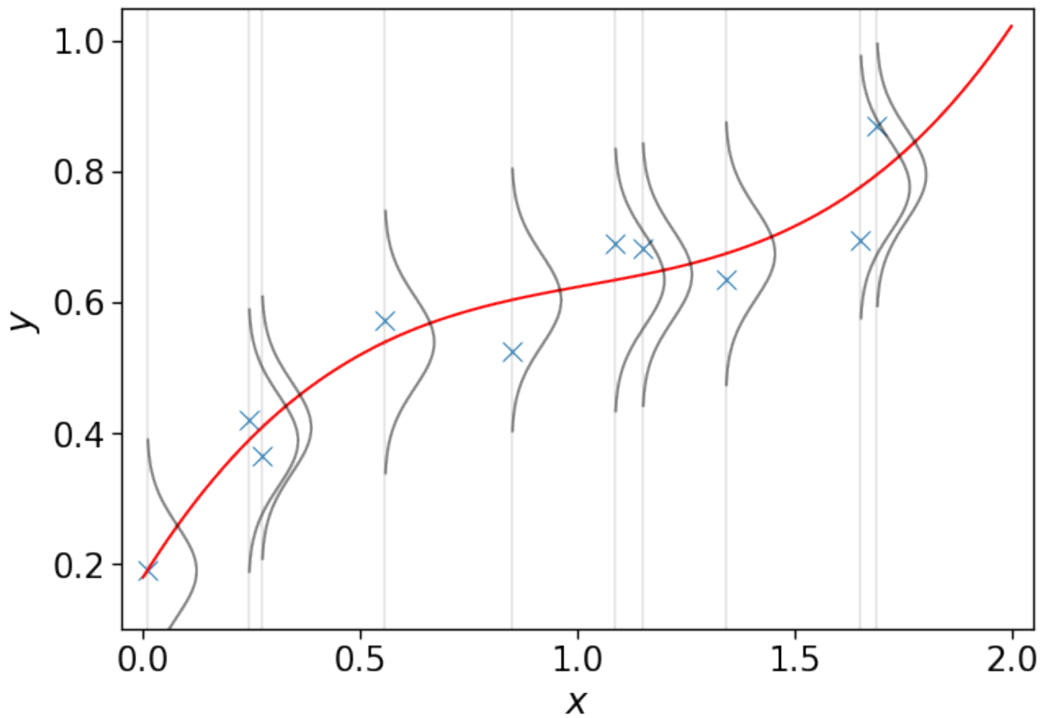
Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \,|\, b)\; p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$



Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \,|\, \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \,|\, \boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

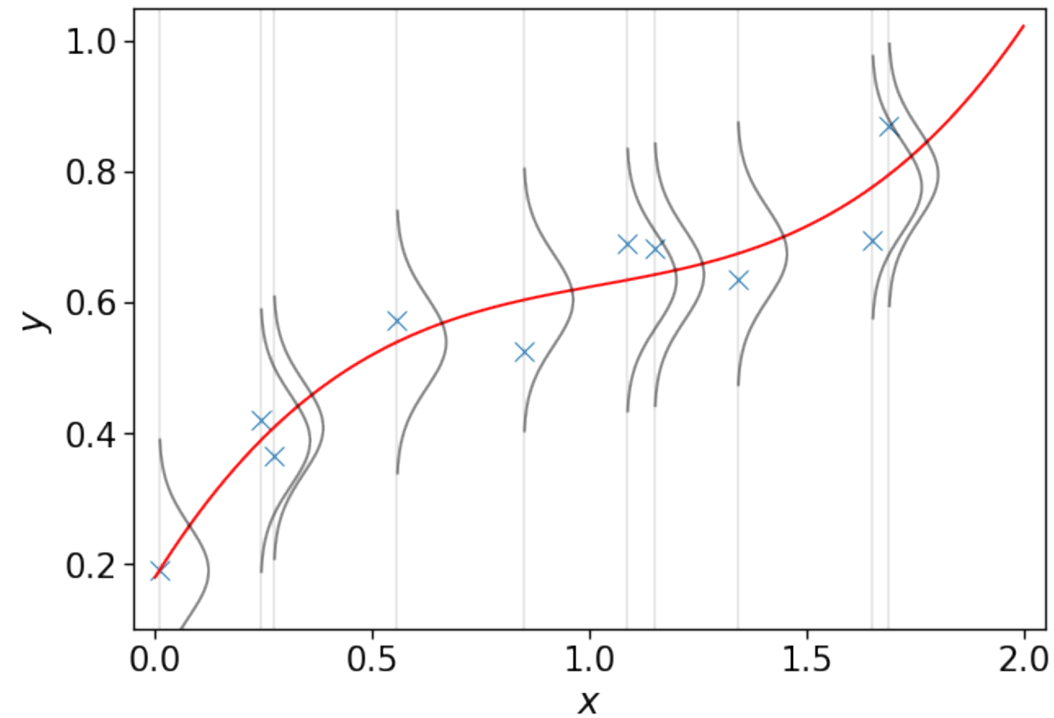
$$p(\boldsymbol{y} \,|\, \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \,|\, \boldsymbol{w}, \sigma^2)$$

Inference

$$p(\boldsymbol{w} \,|\, \boldsymbol{y}) = \frac{p(\boldsymbol{y} \,|\, \boldsymbol{w})\; p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Bayesian Linear Regression

$$p(a) = \int p(a,b)\mathbf{d}b \qquad \& \qquad p(a,b) = p(a\,|\,b)\;p(b)$$



Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$

Model

$$y_n = f_{\mathbf{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n\,|\,\boldsymbol{w},\sigma^2) = \mathscr{N}(y_n\,|\,\boldsymbol{w}^T\boldsymbol{x}_n,\sigma^2)$$

$$p(\boldsymbol{y}\,|\,\boldsymbol{w},\sigma^2) = \prod_{n=1}^{N} p(y_n\,|\,\boldsymbol{w},\sigma^2)$$

Inference

$$p(\boldsymbol{w}\,|\,\boldsymbol{y}) = \frac{p(\boldsymbol{y}\,|\,\boldsymbol{w})\;p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new}\,|\,\boldsymbol{y}) = \int p(y_{new}\,|\,\boldsymbol{w})\;p(\boldsymbol{w}\,|\,\boldsymbol{y})\mathbf{d}\boldsymbol{w}$$

Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\; p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \mid \boldsymbol{w}^T \boldsymbol{x}_n, \sigma^2)$$

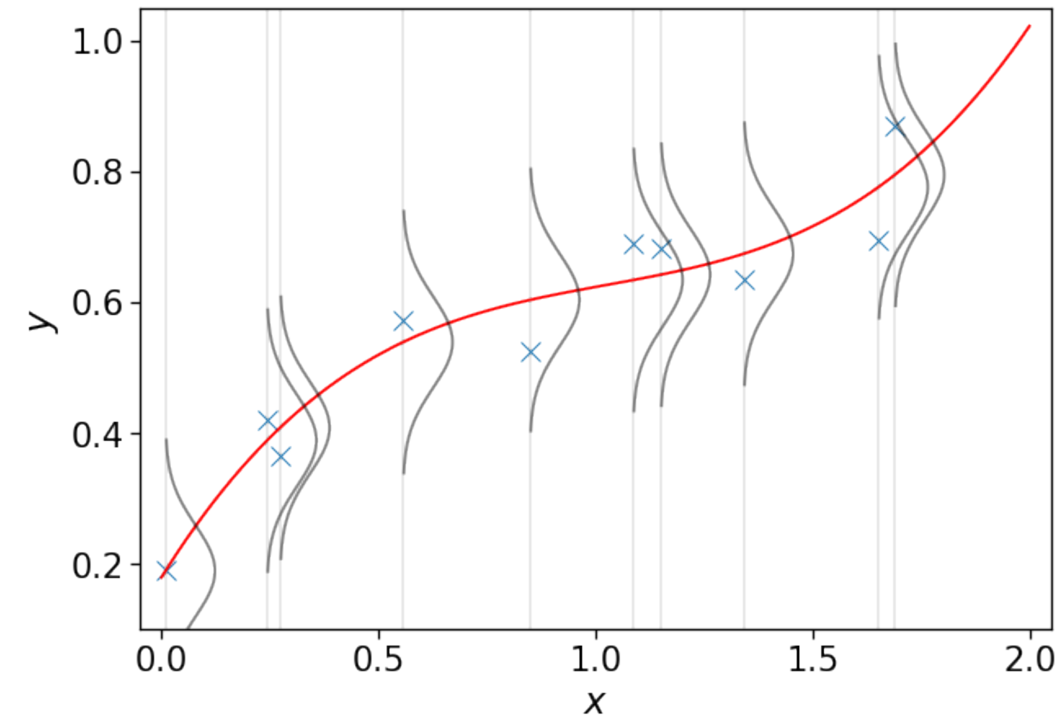
$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

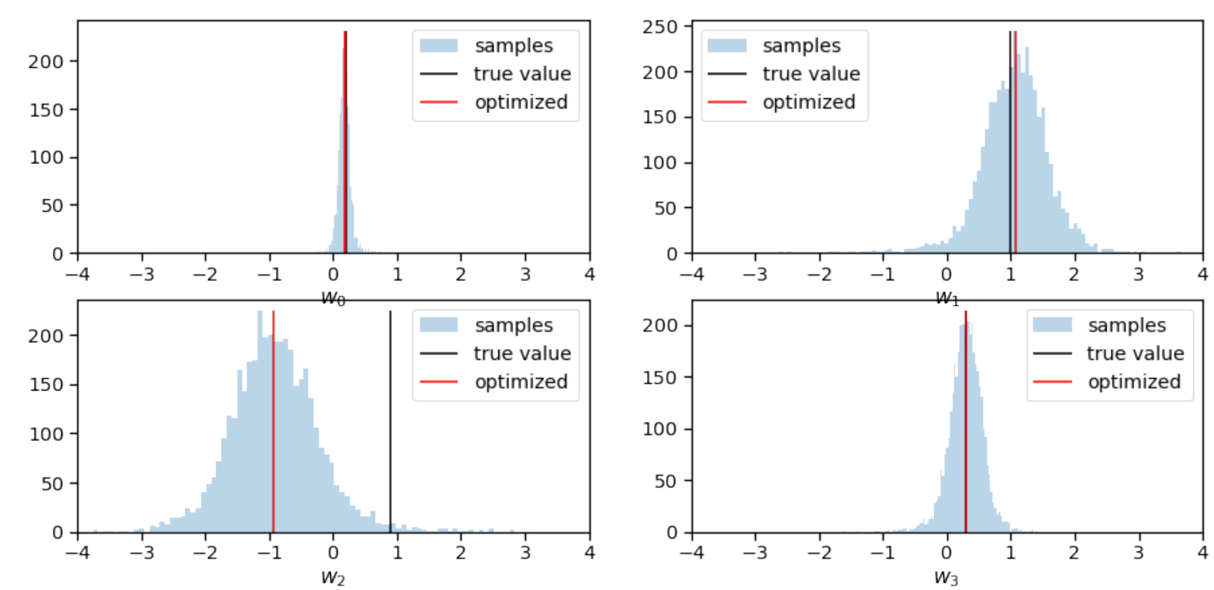
Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\; p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\; p(\boldsymbol{w} \mid \boldsymbol{y})\mathbf{d}\boldsymbol{w}$$





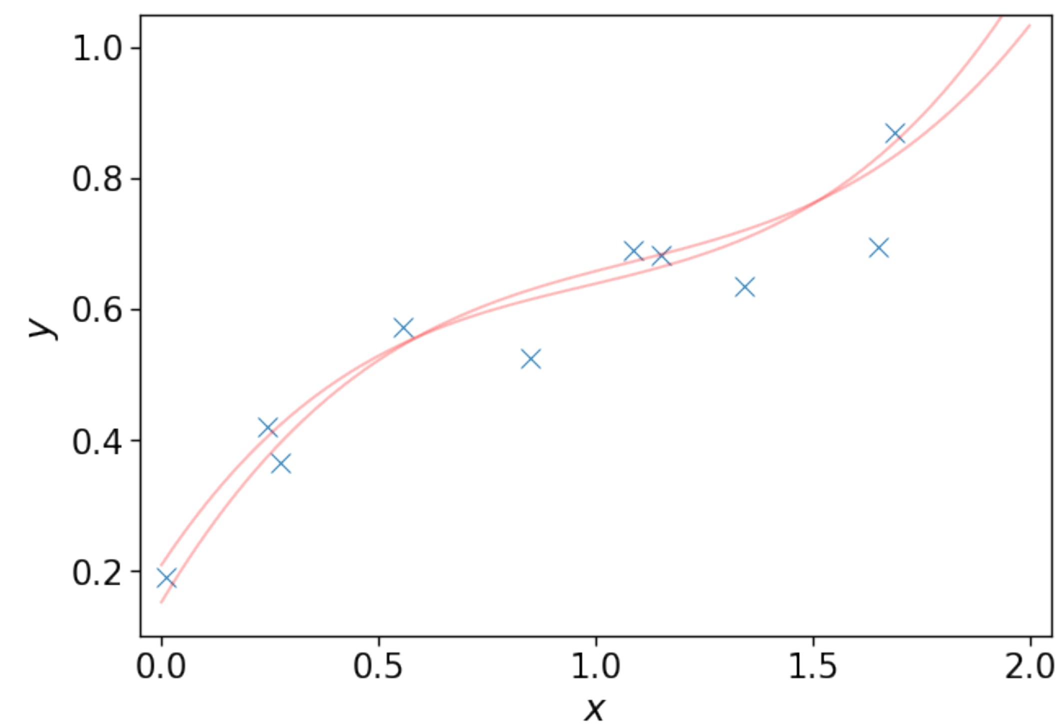
Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\, p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathscr{N}(y_n \mid \boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

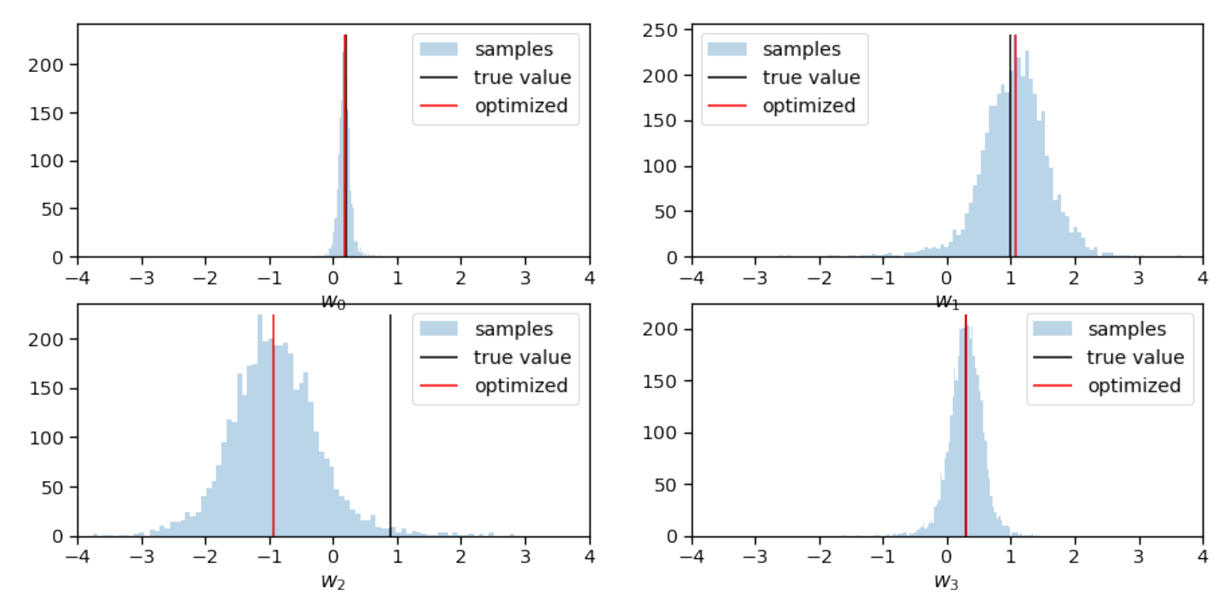
$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\ p(\boldsymbol{w} \mid \boldsymbol{y})\mathbf{d}\boldsymbol{w}$$



Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \mid \boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

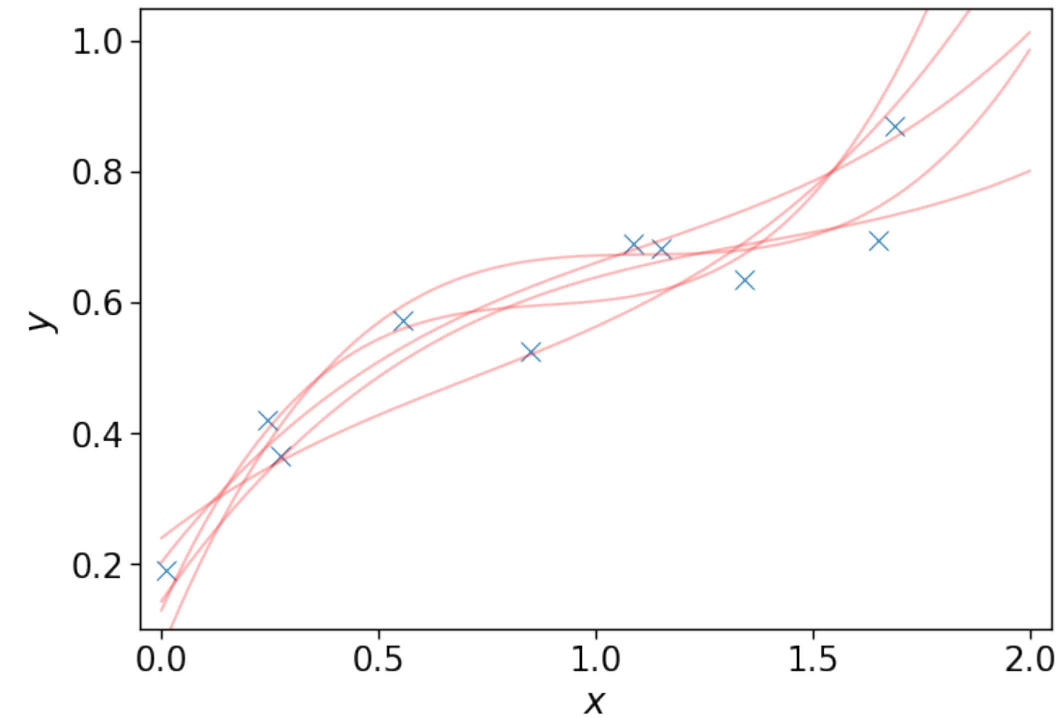
$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

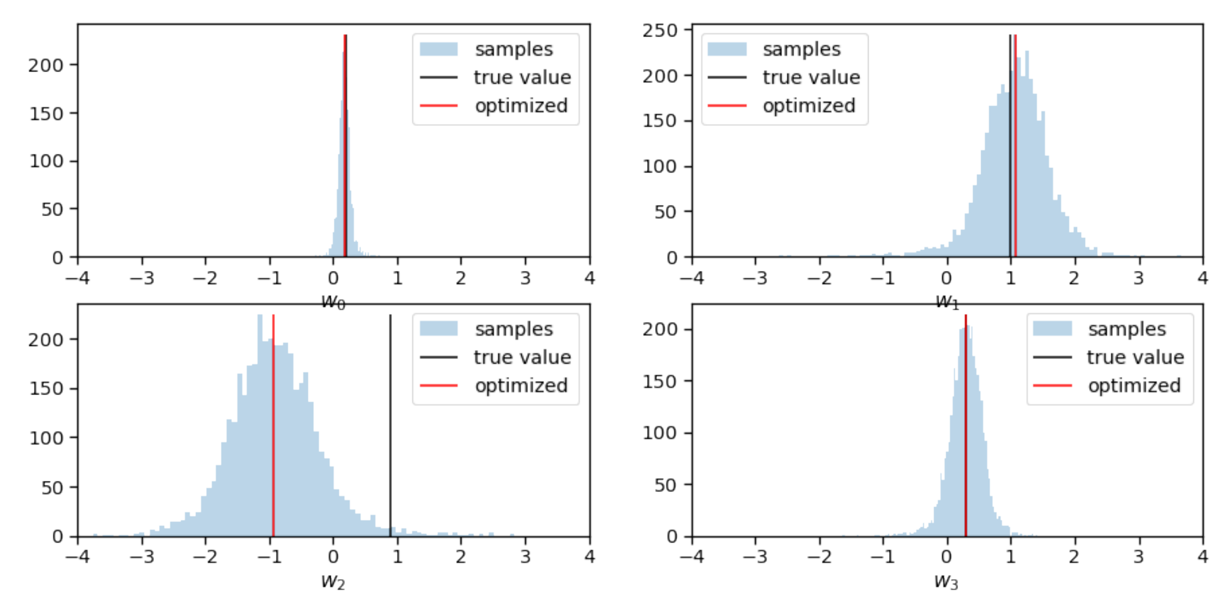
Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\ p(\boldsymbol{w} \mid \boldsymbol{y})\mathbf{d}\boldsymbol{w}$$





Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \mid \boldsymbol{w}^T \boldsymbol{x}_n, \sigma^2)$$

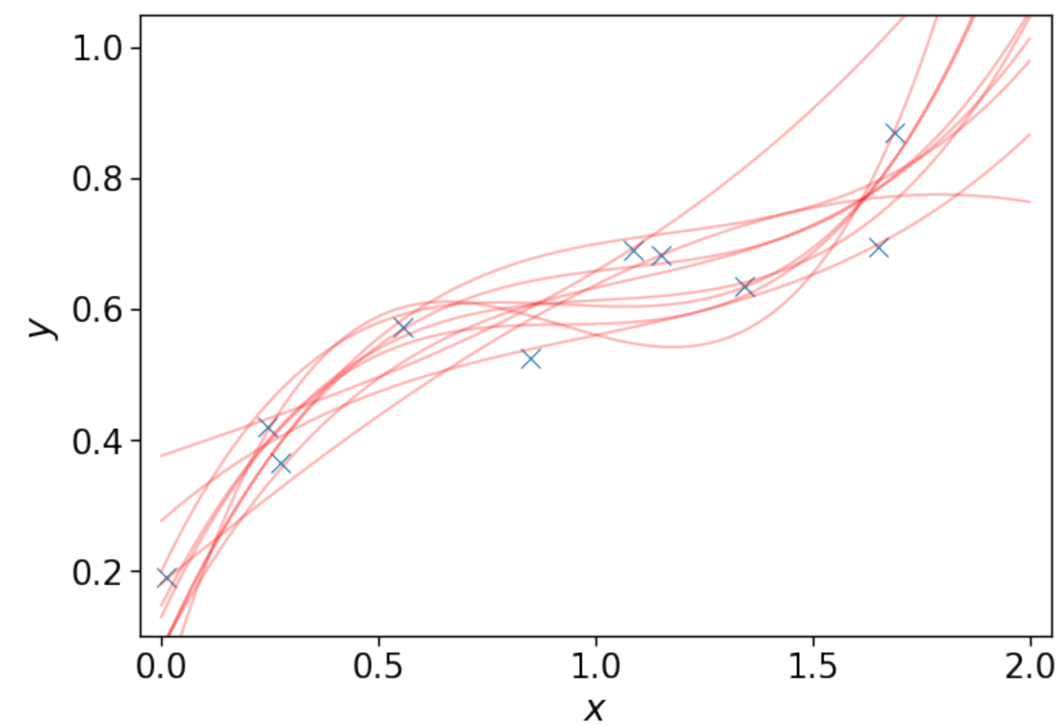
$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

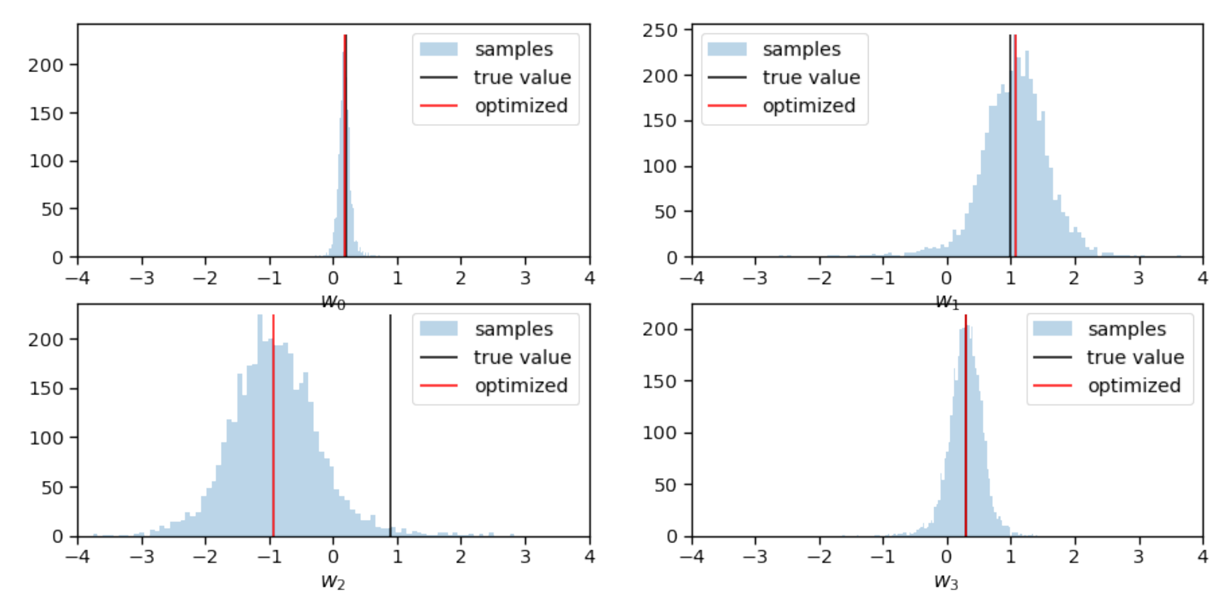
Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\ p(\boldsymbol{w} \mid \boldsymbol{y})\mathbf{d}\boldsymbol{w}$$





Bayesian Linear Regression

$$p(a) = \int p(a, b) \mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \mid \boldsymbol{w}^T \boldsymbol{x}_n, \sigma^2)$$

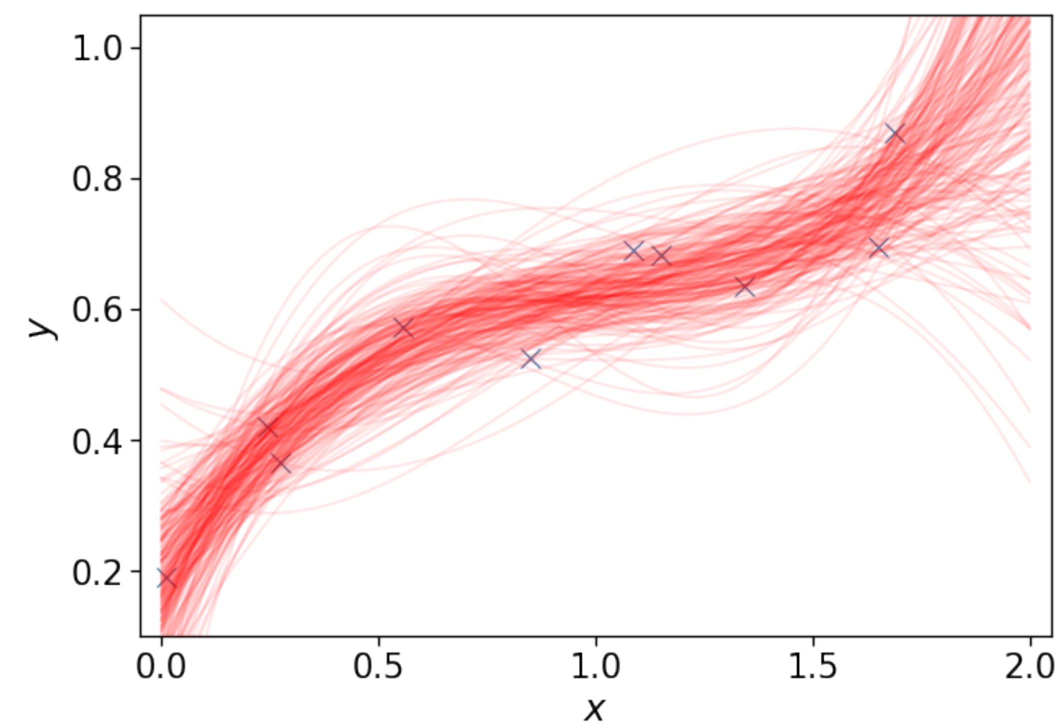
$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

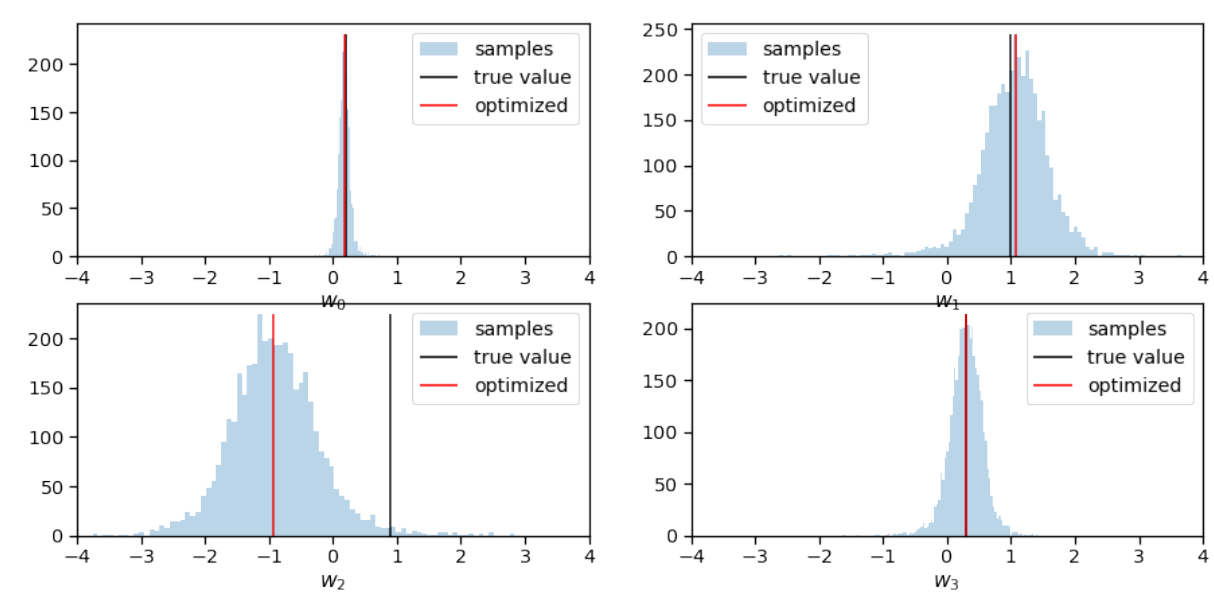
Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\ p(\boldsymbol{w} \mid \boldsymbol{y}) \mathbf{d}\boldsymbol{w}$$





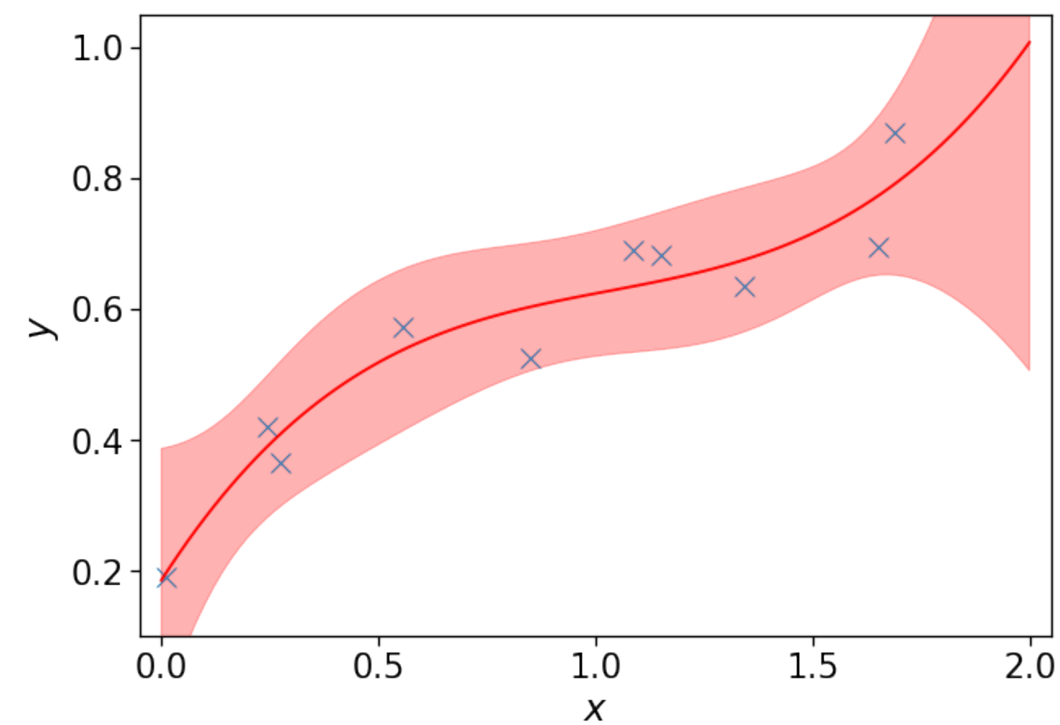
Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\ p(b)$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$



Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \mid \boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\ p(\boldsymbol{w} \mid \boldsymbol{y})\mathbf{d}\boldsymbol{w}$$

Bayesian Linear Regression

$$p(a) = \int p(a, b)\mathbf{d}b \qquad \& \qquad p(a, b) = p(a \mid b)\; p(b)$$

Observed Data

$$\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\; y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T \boldsymbol{x}_n + \epsilon_n$$

$$p(y_n \mid \boldsymbol{w}, \sigma^2) = \mathcal{N}(y_n \mid \boldsymbol{w}^T \boldsymbol{x}_n, \sigma^2)$$

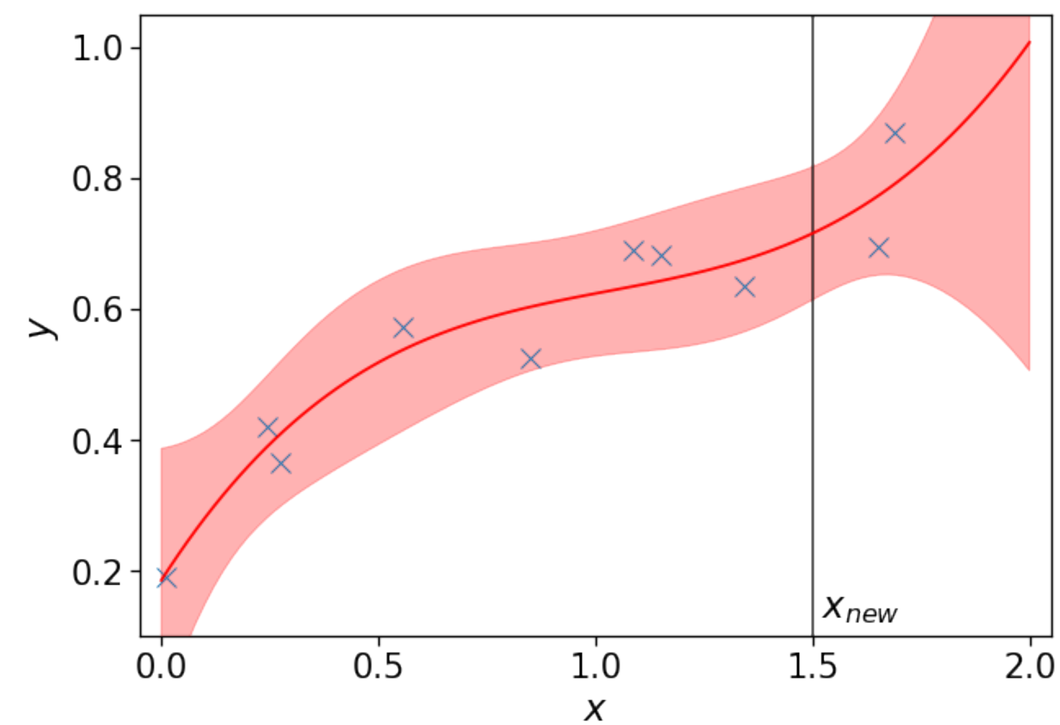
$$p(\boldsymbol{y} \mid \boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n \mid \boldsymbol{w}, \sigma^2)$$

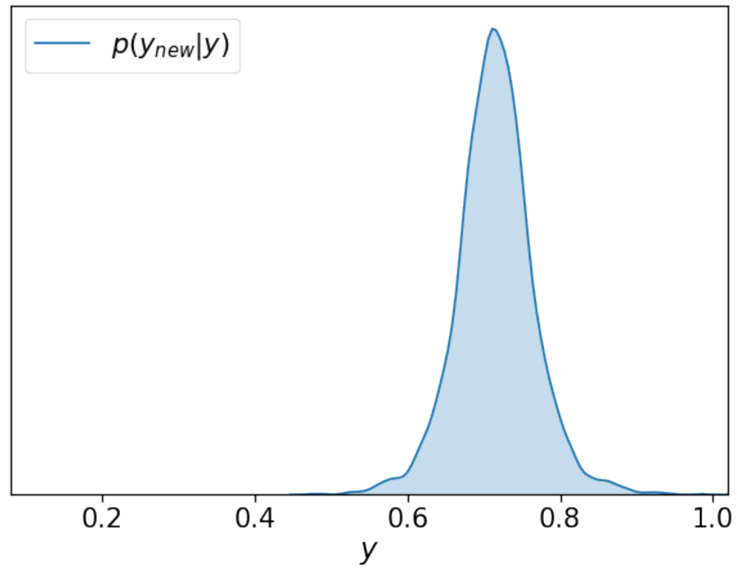
Inference

$$p(\boldsymbol{w} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{w})\; p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \mid \boldsymbol{y}) = \int p(y_{new} \mid \boldsymbol{w})\; p(\boldsymbol{w} \mid \boldsymbol{y})\mathbf{d}\boldsymbol{w}$$





Comparison

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N}\left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2$$

Fit

$$\boldsymbol{w}* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$$

Prediction

$$y_{new} = f_{\boldsymbol{w}*}(\boldsymbol{x}_{new})$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$$

Model

$$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$$

$$p(\boldsymbol{y}\,|\,\boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N}\mathcal{N}(y_n\,|\,\boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$$

Inference

$$p(\boldsymbol{w}\,|\,\boldsymbol{y}) = \frac{p(\boldsymbol{y}\,|\,\boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new}\,|\,\boldsymbol{y}) = \int p(y_{new}\,|\,\boldsymbol{w})\ p(\boldsymbol{w}\,|\,\boldsymbol{y})\mathbf{d}\boldsymbol{w}$$

Regularized Version

Observed Data	$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$ $\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$	**Observed Data**	$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$ $\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \mathbb{R}$
Model	$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$ $E_{\mathscr{D}}(\boldsymbol{w}) = \sum_{n=1}^{N} \left(y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right)^2$ $\boxed{+\lambda\|\boldsymbol{w}\|_2^2}$	**Model**	$y_n = f_{\boldsymbol{w}}(\boldsymbol{x}_n) + \epsilon_n = \boldsymbol{w}^T\boldsymbol{x}_n + \epsilon_n$ $p(\boldsymbol{y}\,\vert\,\boldsymbol{w}, \sigma^2) = \prod_{n=1}^{N} \mathscr{N}(y_n\,\vert\,\boldsymbol{w}^T\boldsymbol{x}_n, \sigma^2)$ $\boxed{p(\boldsymbol{w}) = \mathscr{N}(\boldsymbol{w}\,\vert\,\boldsymbol{0}, \boldsymbol{I})}$
Fit	$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$	**Inference**	$p(\boldsymbol{w}\,\vert\,\boldsymbol{y}) = \frac{p(\boldsymbol{y}\,\vert\,\boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$
Prediction	$y_{new} = f_{\boldsymbol{w}^*}(\boldsymbol{x}_{new})$	**Prediction**	$p(y_{new}\,\vert\,\boldsymbol{y}) = \int p(y_{new}\,\vert\,\boldsymbol{w})\ p(\boldsymbol{w}\,\vert\,\boldsymbol{y})\mathbf{d}\boldsymbol{w}$

Classification

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \{0,1\}$$

Model

$$p(y_n = 1) = \sigma\left(f_{\boldsymbol{w}}(\boldsymbol{x}_n)\right) = \sigma\left(\boldsymbol{w}^T \boldsymbol{x}_n\right)$$

$$E_{\mathscr{D}}(\boldsymbol{w}) = -\sum_{n=1}^{N} \Big[y_n \log(\sigma(\boldsymbol{w}^T \boldsymbol{x}_n)) + (1 - y_n)\log(1 - \sigma(\boldsymbol{w}^T \boldsymbol{x}_n)\Big]$$

Fit

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} E_{\mathscr{D}}(\boldsymbol{w})$$

Prediction

$$y_{new} = f_{\boldsymbol{w}^*}(\boldsymbol{x}_{new})$$

Observed Data

$$\mathscr{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$$

$$\boldsymbol{x}_n \in \mathbb{R}^D,\ y_n \in \{0,1\}$$

Model

$$p(y_n = 1) = \sigma\left(f_{\boldsymbol{w}}(\boldsymbol{x}_n)\right) = \sigma\left(\boldsymbol{w}^T \boldsymbol{x}_n\right)$$

$$p(\boldsymbol{y} \,|\, \boldsymbol{w}) = \prod_{n=1}^{N} \sigma(\boldsymbol{w}^T \boldsymbol{x}_n)^{y_n}\ (1 - \sigma(\boldsymbol{w}^T \boldsymbol{x}_n))^{1-y_n}$$

Inference

$$p(\boldsymbol{w} \,|\, \boldsymbol{y}) = \frac{p(\boldsymbol{y} \,|\, \boldsymbol{w})\ p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

Prediction

$$p(y_{new} \,|\, \boldsymbol{y}) = \int p(y_{new} \,|\, \boldsymbol{w})\ p(\boldsymbol{w} \,|\, \boldsymbol{y}) \mathbf{d}\boldsymbol{w}$$

Asymptotic Certainty

- If $\mathscr{D}_N = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$ are generated by some true $\boldsymbol{w}^*$, as long as $p(\boldsymbol{w}^*) > 0$, we have

$$\lim_{N \to \infty} p(\boldsymbol{w} \mid \mathscr{D}_N) = \delta(\boldsymbol{w} - \boldsymbol{w}^*)$$

- Unrealisable case: Data generated by some $p^*(\mathscr{D})$ but cannot be modelled by any $\boldsymbol{w}$, then

$$\lim_{N \to \infty} p(\boldsymbol{w} \mid \mathscr{D}_N) = \delta(\boldsymbol{w} - \hat{\boldsymbol{w}})$$

where $\hat{\boldsymbol{w}}$ minimizes KL $\left[p^*(\mathscr{D}) \| p(\mathscr{D} \mid \boldsymbol{w})\right]$

Summary of Bayesian ML

$$p(x) = \sum_y p(x, y) \qquad \& \qquad p(x, y) = p(y \mid x)\ p(x)$$

- Inference

$$p(w \mid \mathscr{D}) = \frac{p(\mathscr{D} \mid w)\ p(w)}{p(\mathscr{D})}$$

- Prediction

$$p(y \mid \mathscr{D}) = \int p(y \mid w)\ p(w \mid \mathscr{D}) \mathbf{d}w$$

Summary of Bayesian ML

$$p(x) = \sum_y p(x, y) \qquad \& \qquad p(x, y) = p(y \mid x)\ p(x)$$

- Inference $$p(w \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid w)\ p(w)}{p(\mathcal{D})}$$

- Prediction $$p(y \mid \mathcal{D}) = \int p(y \mid w)\ p(w \mid \mathcal{D}) \mathbf{d}w$$

$m \in \{1,\ldots,M\}$

Summary of Bayesian ML

$$p(x) = \sum_y p(x, y) \qquad \& \qquad p(x, y) = p(y \mid x)\ p(x)$$

- Inference $$p(w \mid \mathscr{D}, m) = \frac{p(\mathscr{D} \mid w, m)\ p(w \mid m)}{p(\mathscr{D} \mid m)}$$

- Prediction $$p(y \mid \mathscr{D}, m) = \int p(y \mid w, m)\ p(w \mid \mathscr{D}, m)\mathbf{d}w$$

$m \in \{1,\ldots,M\}$

- Model Comparison $$p(m \mid \mathscr{D}) = \frac{p(\mathscr{D} \mid m)p(m)}{p(\mathscr{D})}$$

Thanks for your attention!

Supervisor: Prof. Dr. Nils Bertschinger
Funder: Dr. h. c. Helmut O. Maucher