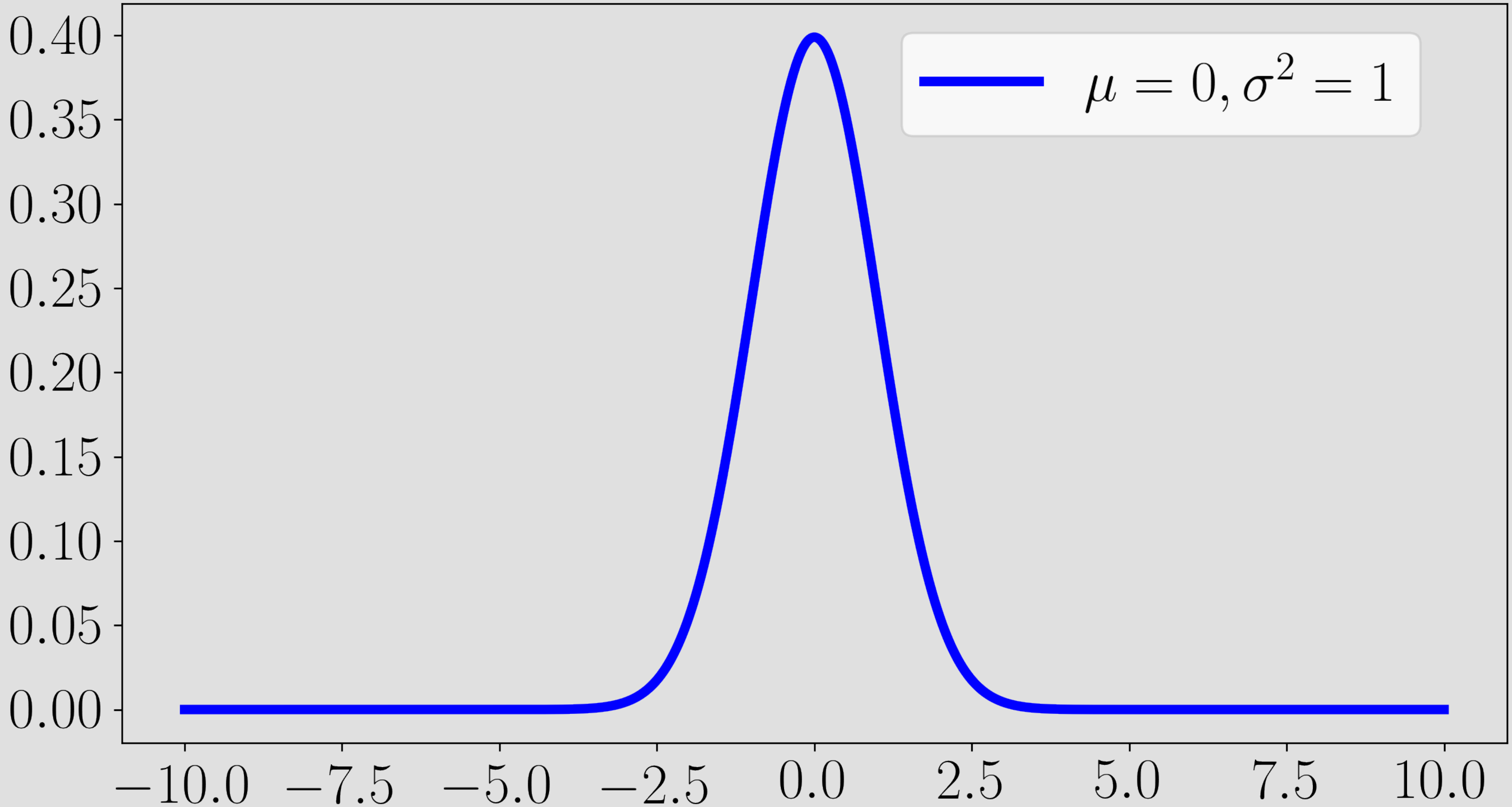
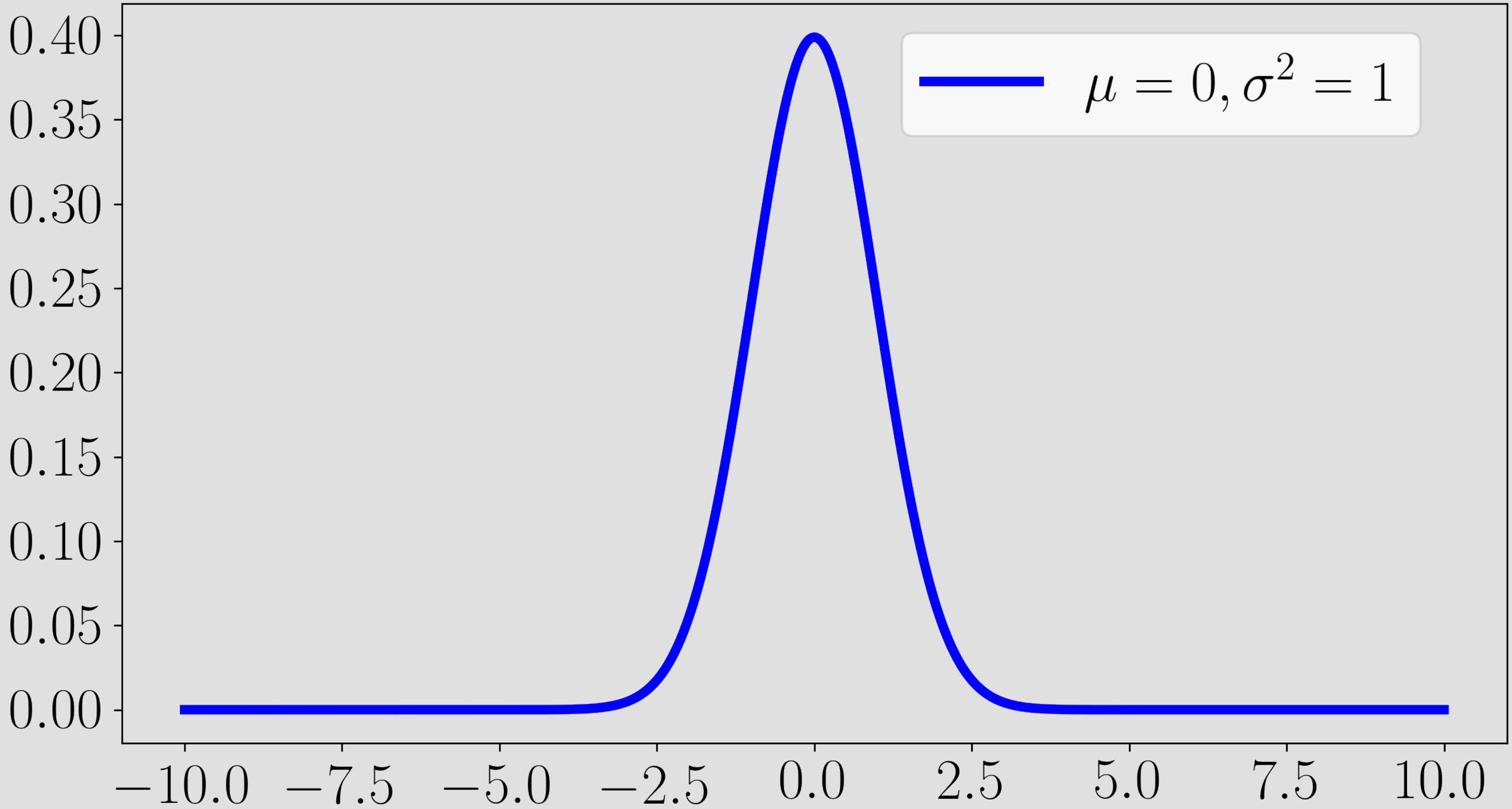
Example: linear regression

Univariate normal



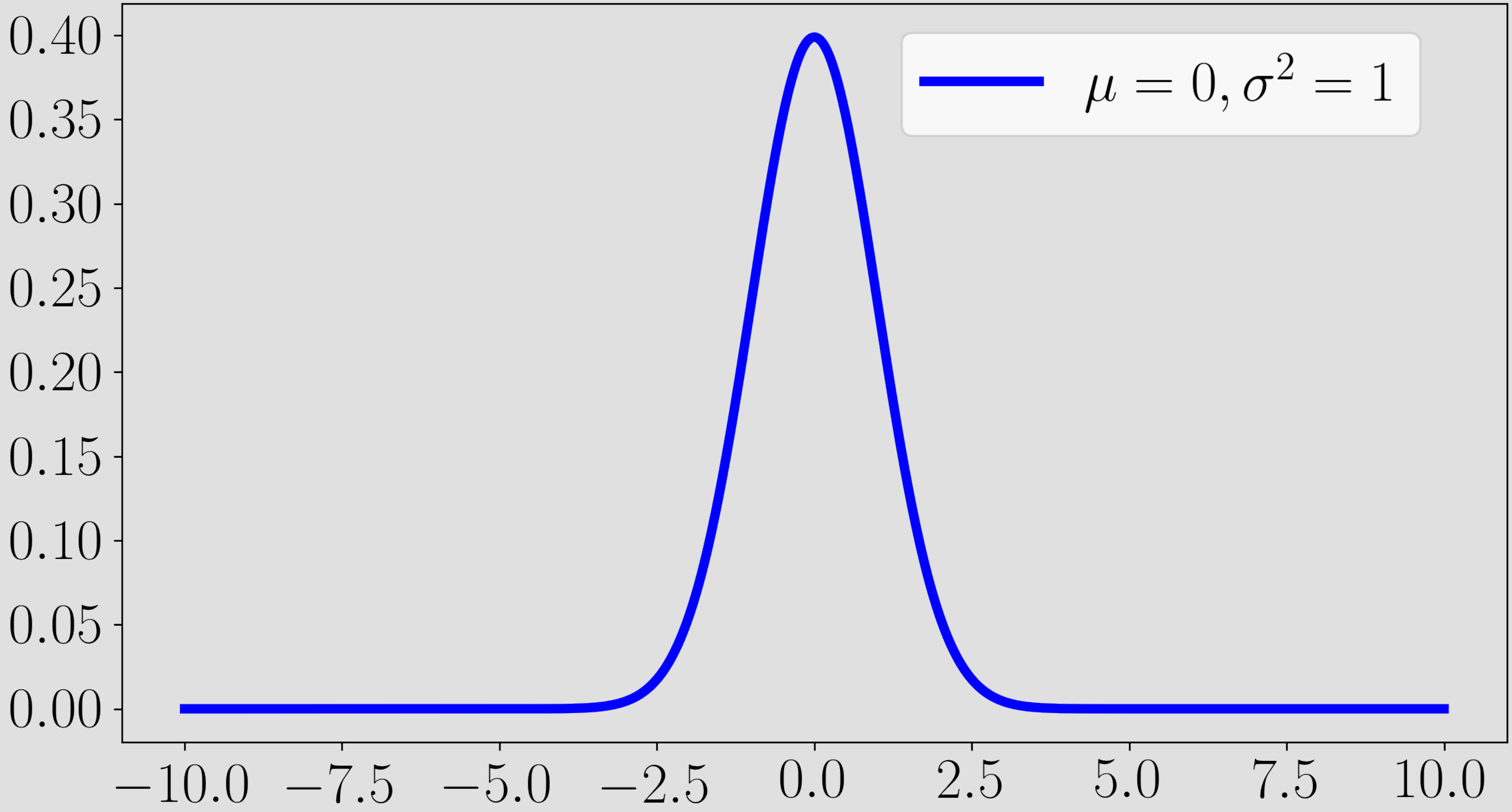
$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Univariate normal



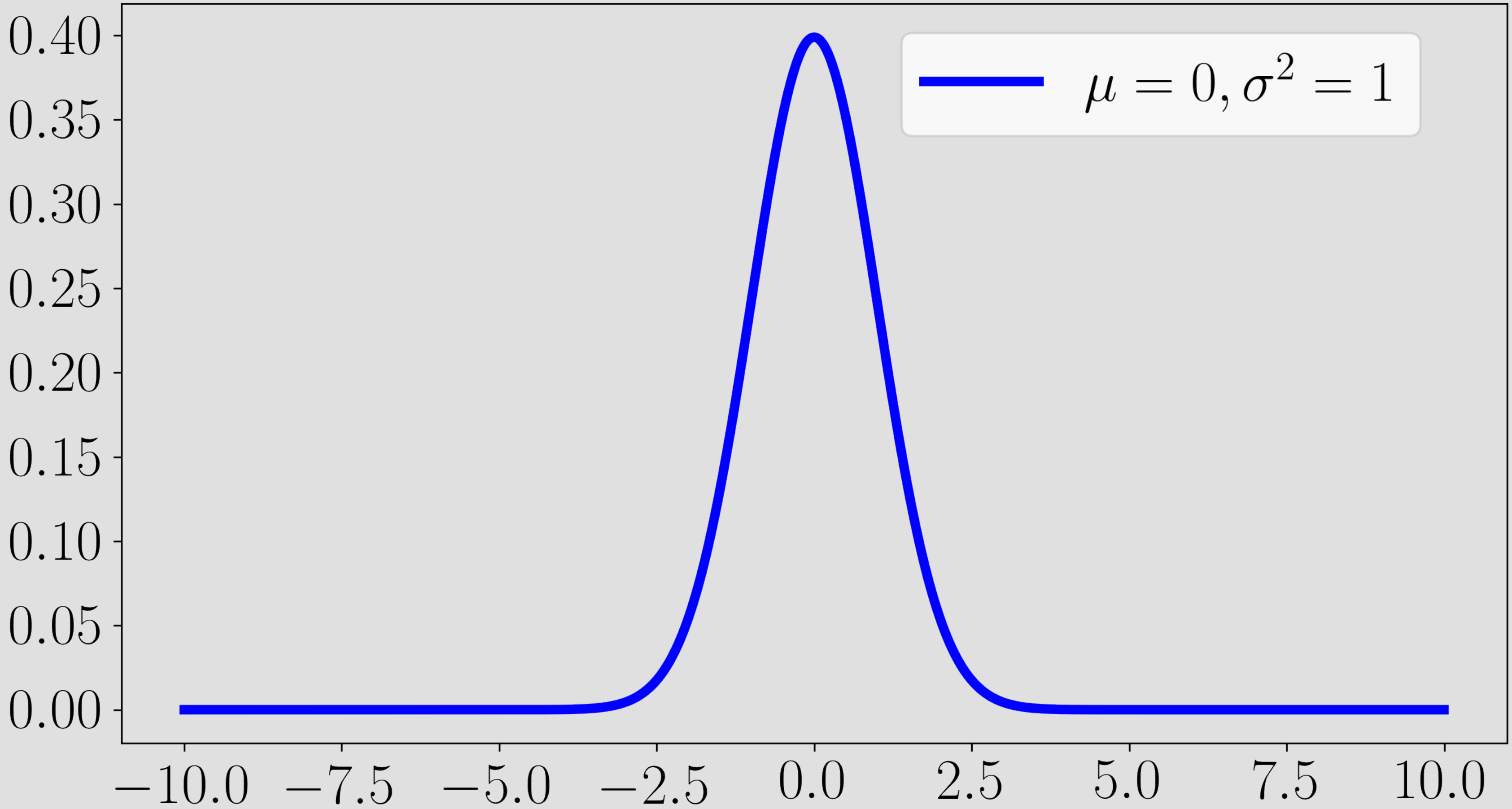
$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Univariate normal



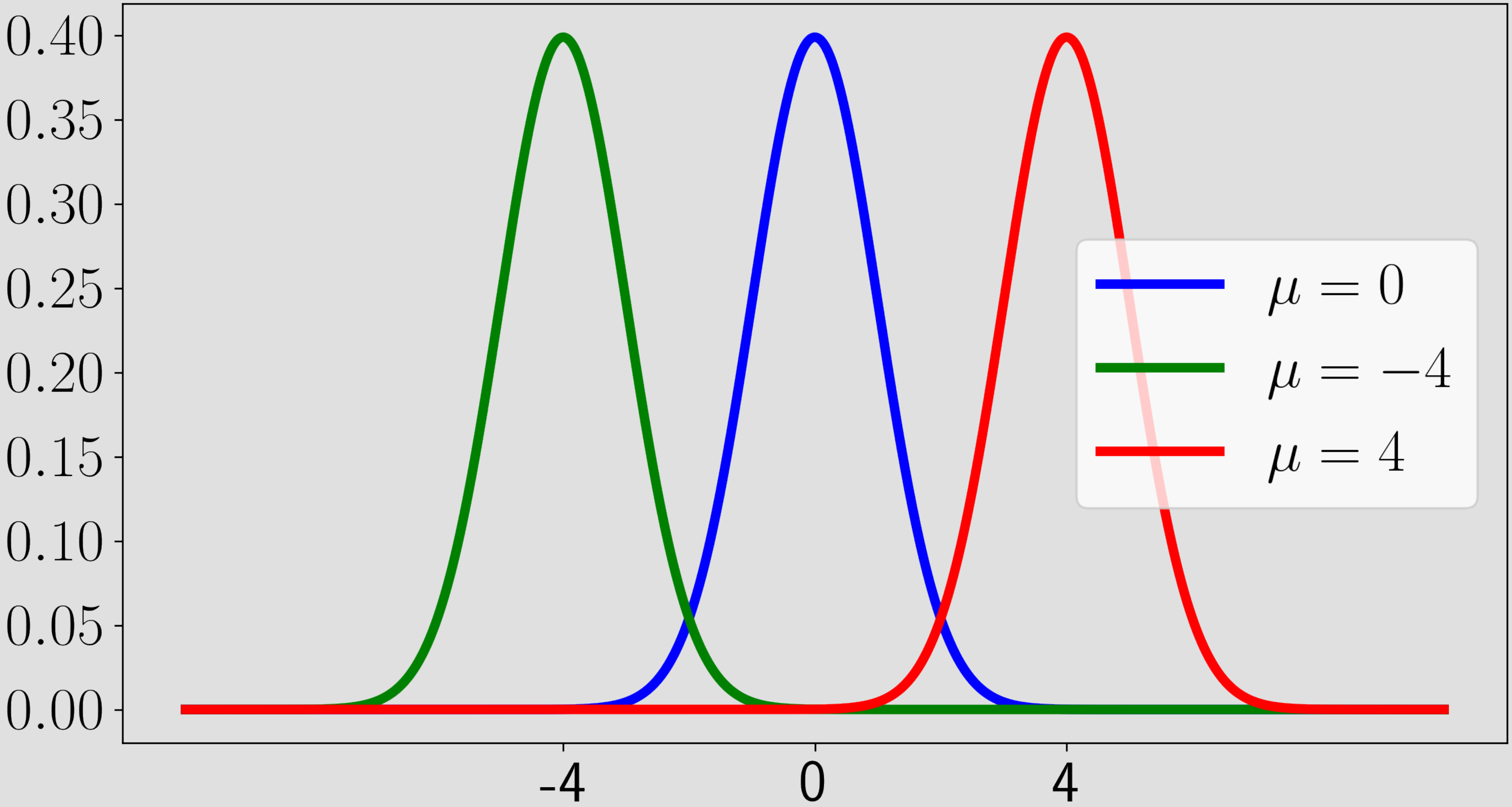
$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Univariate normal



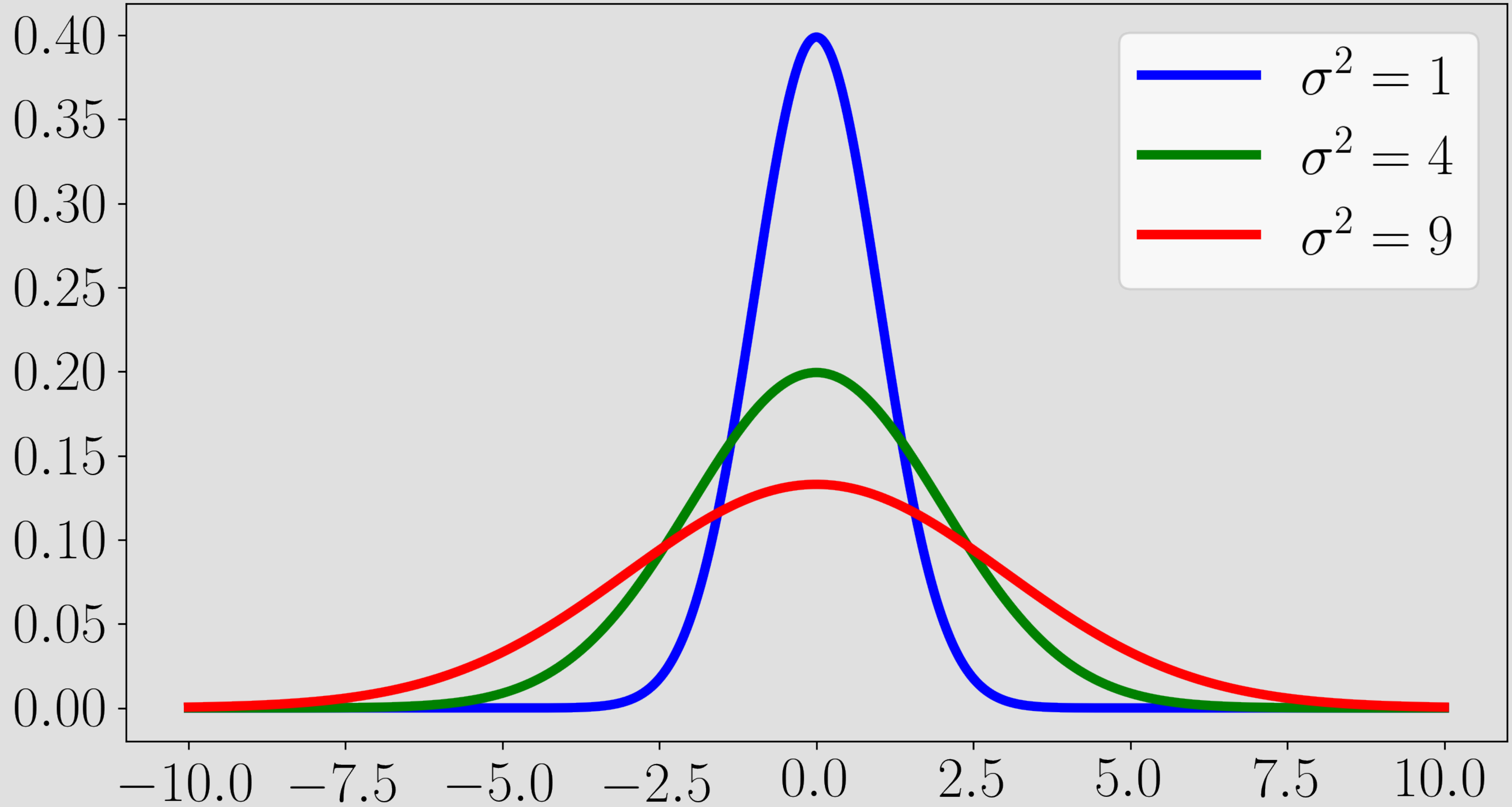
$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Univariate normal: mean



$$\mathbb{E}X = \mu$$

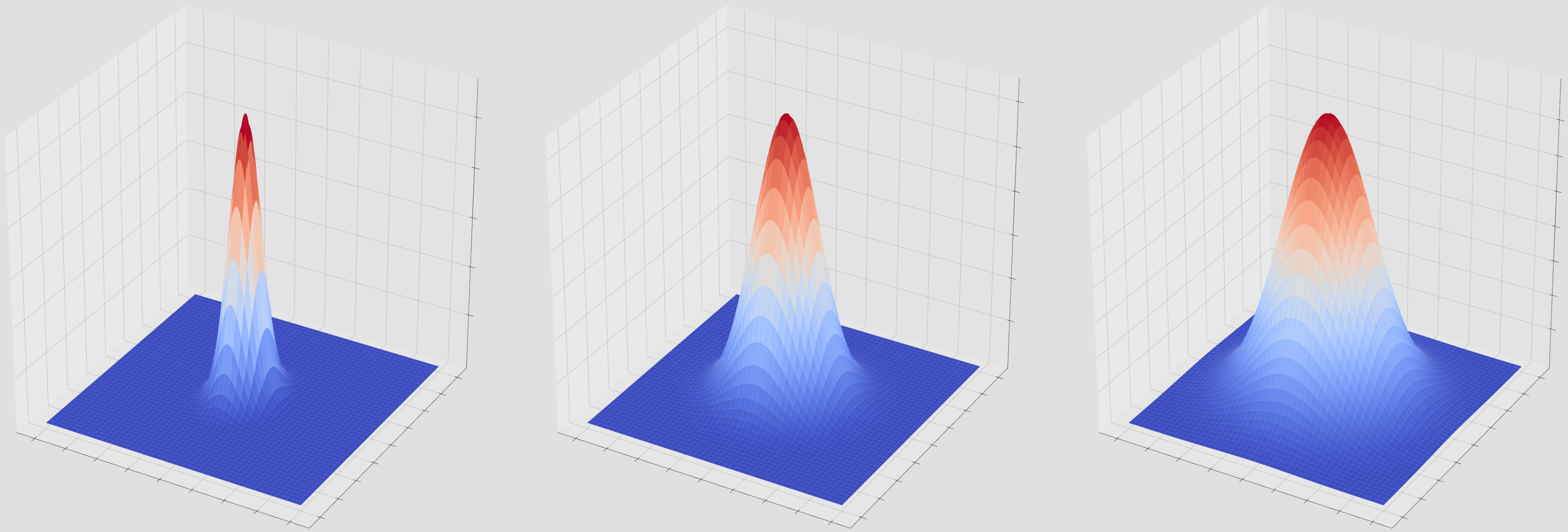
Univariate normal: variance



$$\mathrm{Var}[X] = \sigma^2$$

Multivariate normal

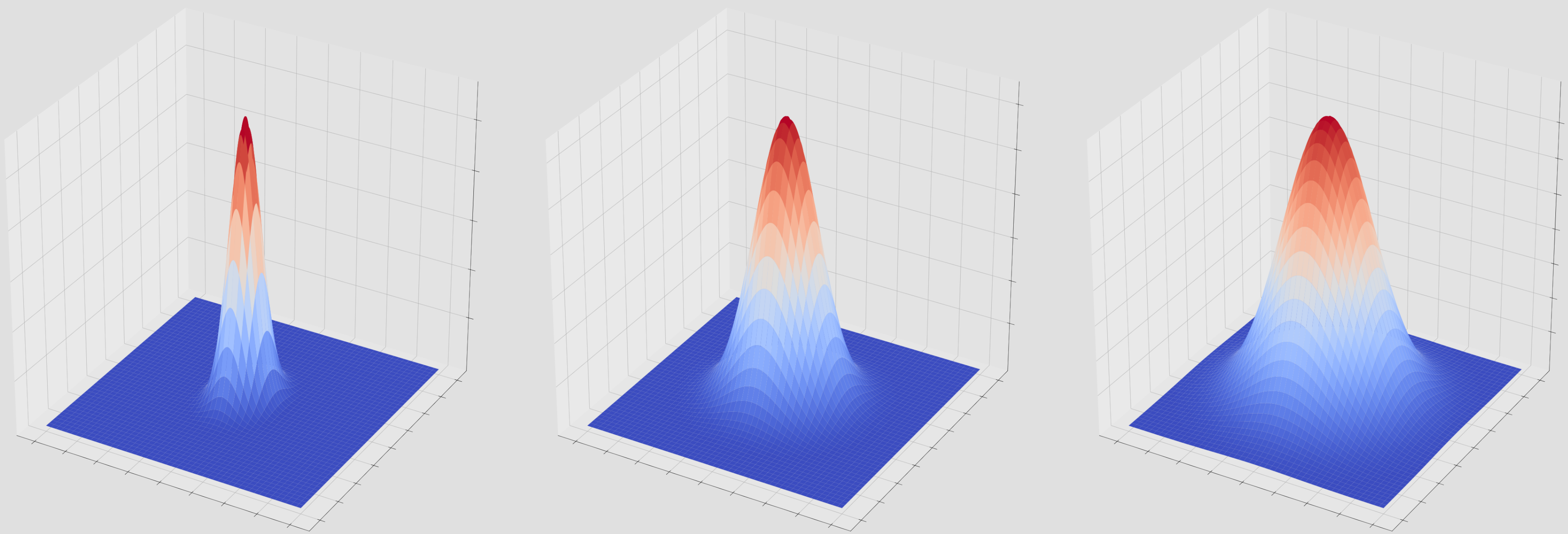
$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$



Multivariate normal

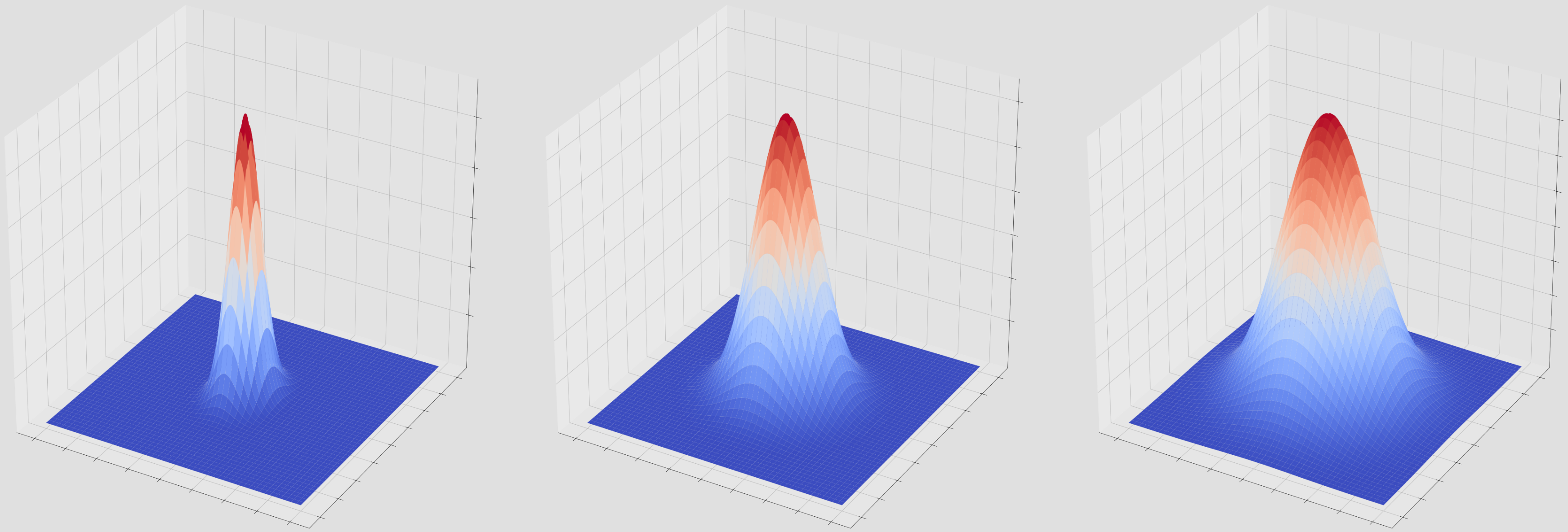
$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$$\mathbb{E}X = \mu \quad \mathrm{Cov}[X] = \Sigma$$



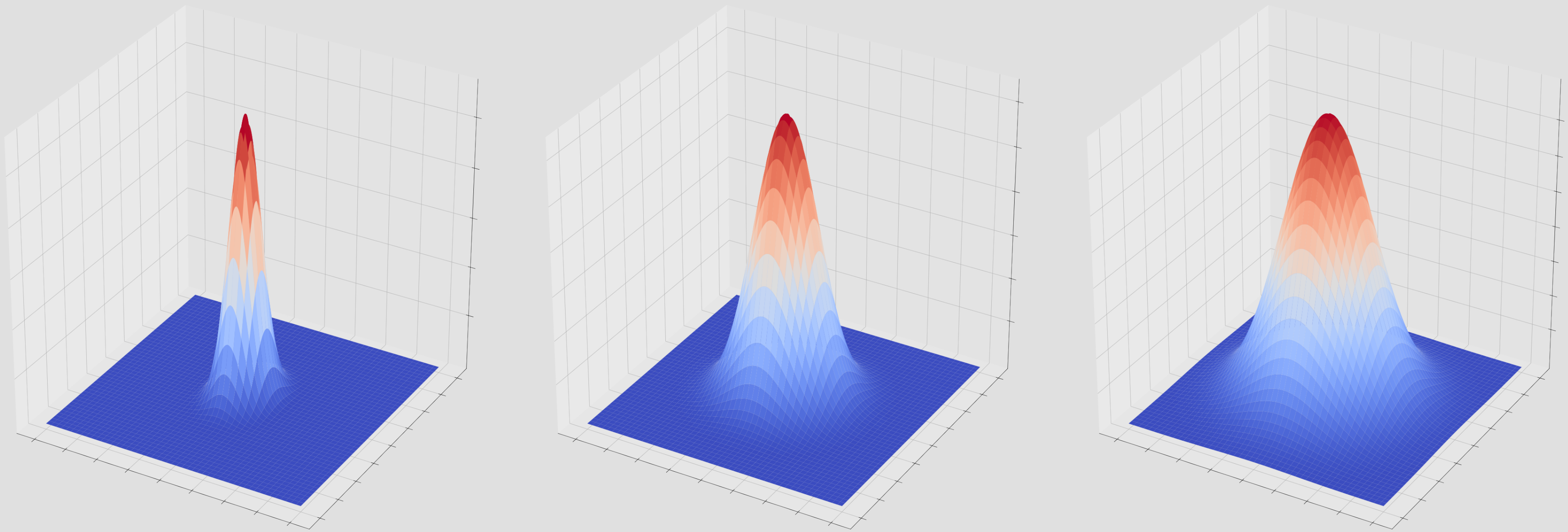
Multivariate normal

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$



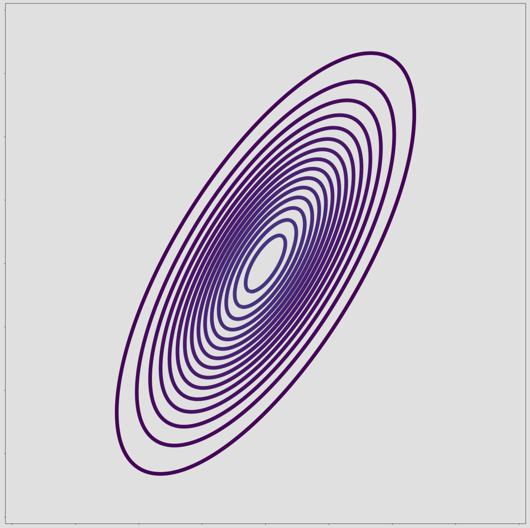
Multivariate normal

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$



Multivariate normal

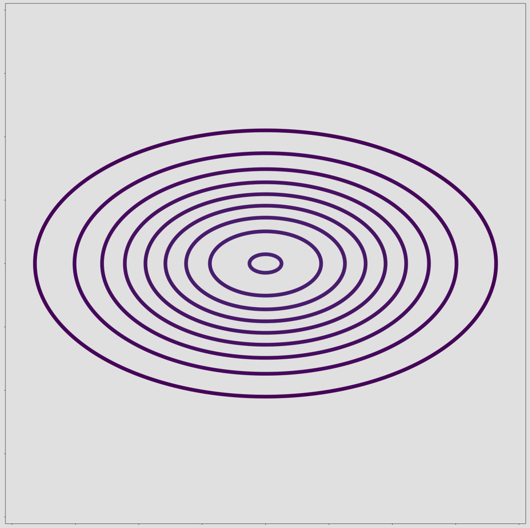
$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$



Full

Parameters: $\frac{D(D+1)}{2}$

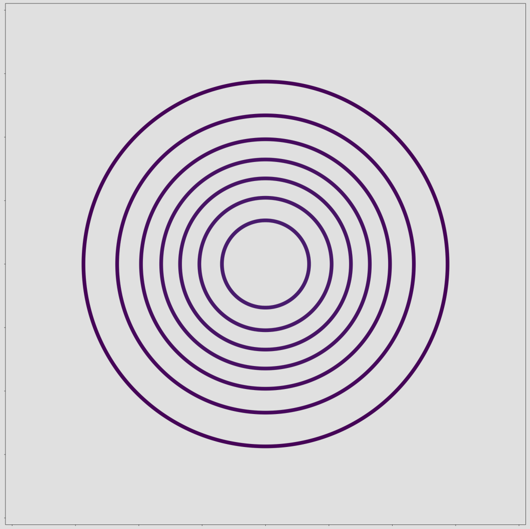
$$\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$



Diagonal

Parameters: D

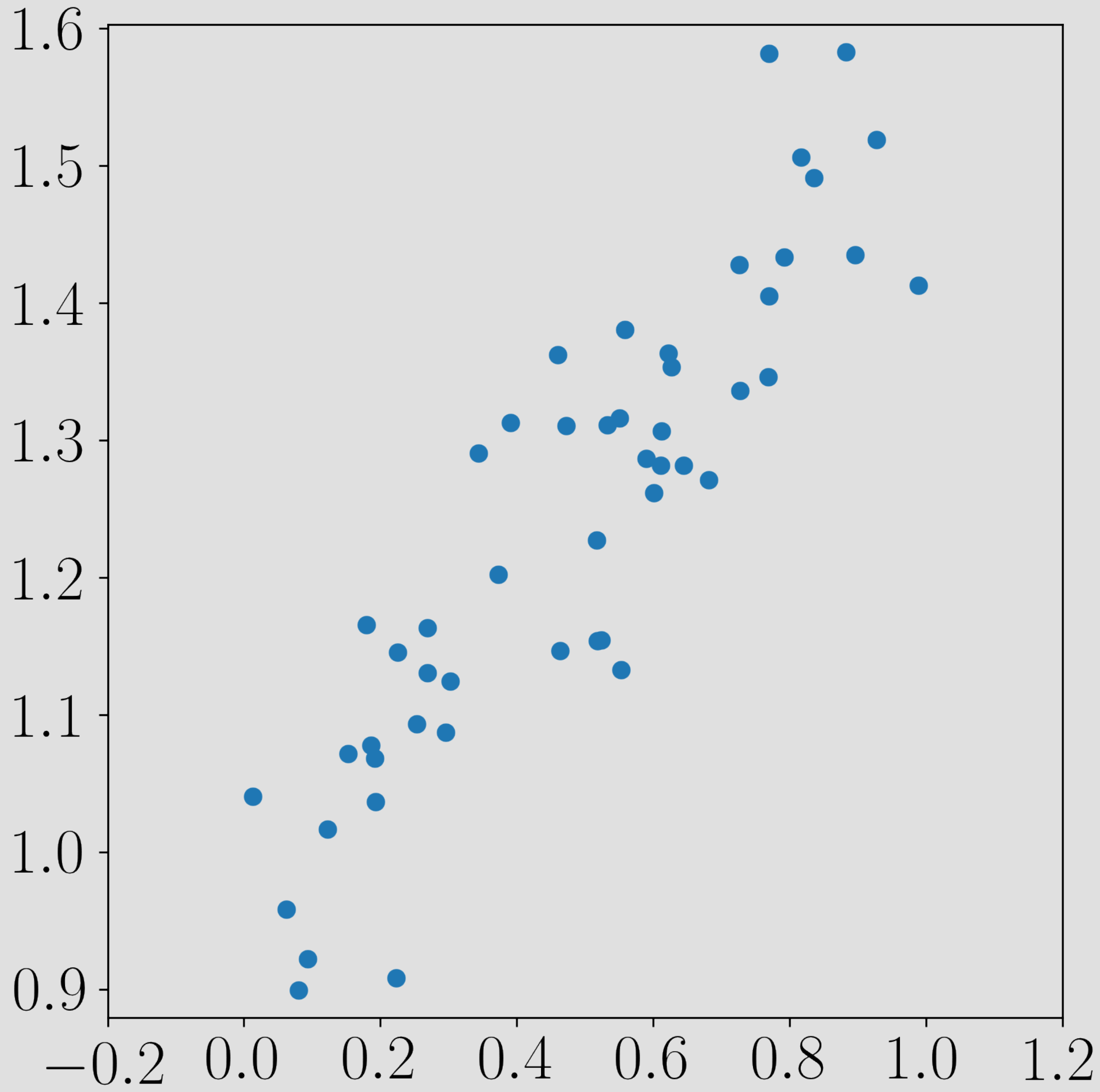
$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$



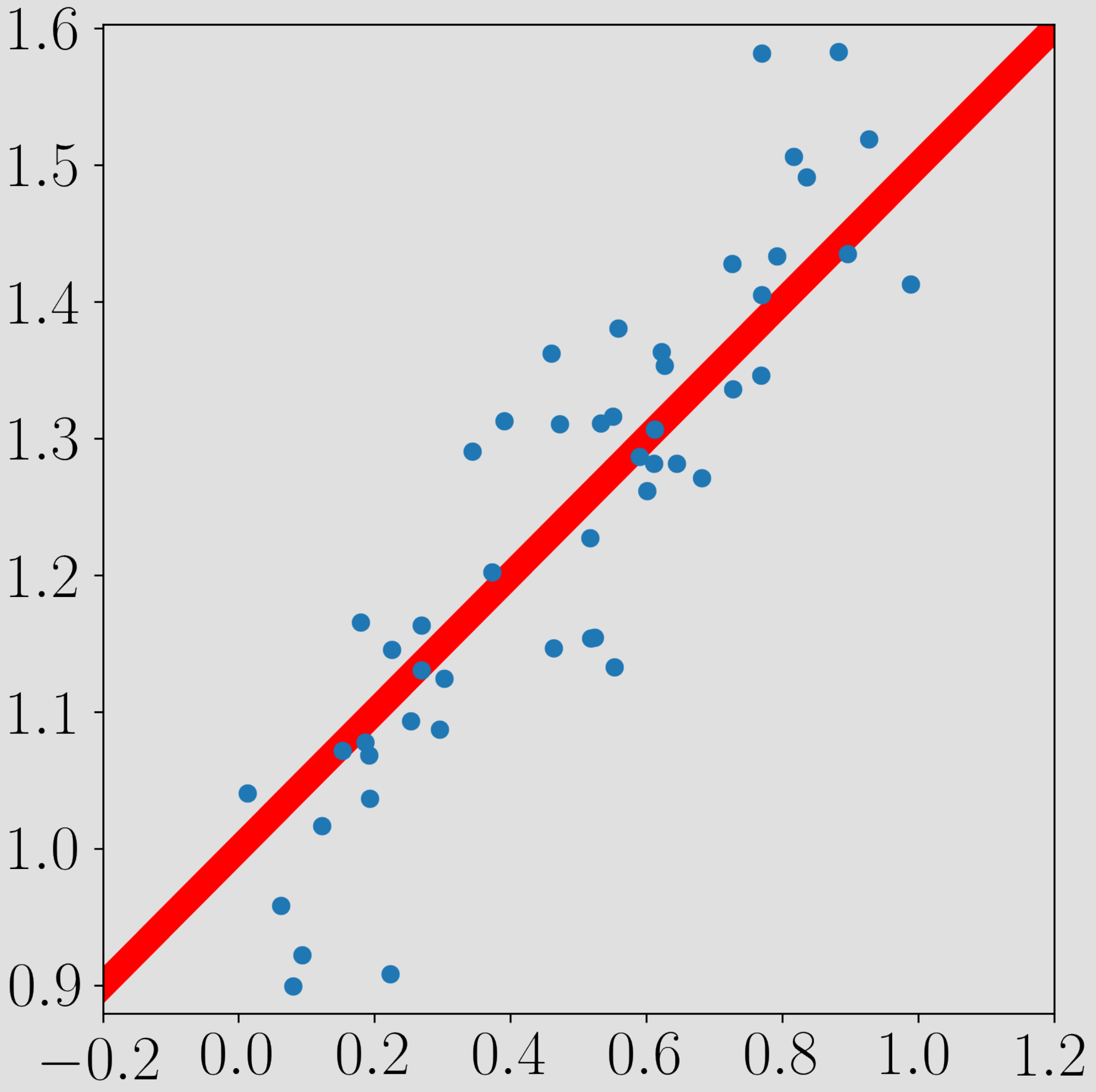
Spherical

Parameters: 1

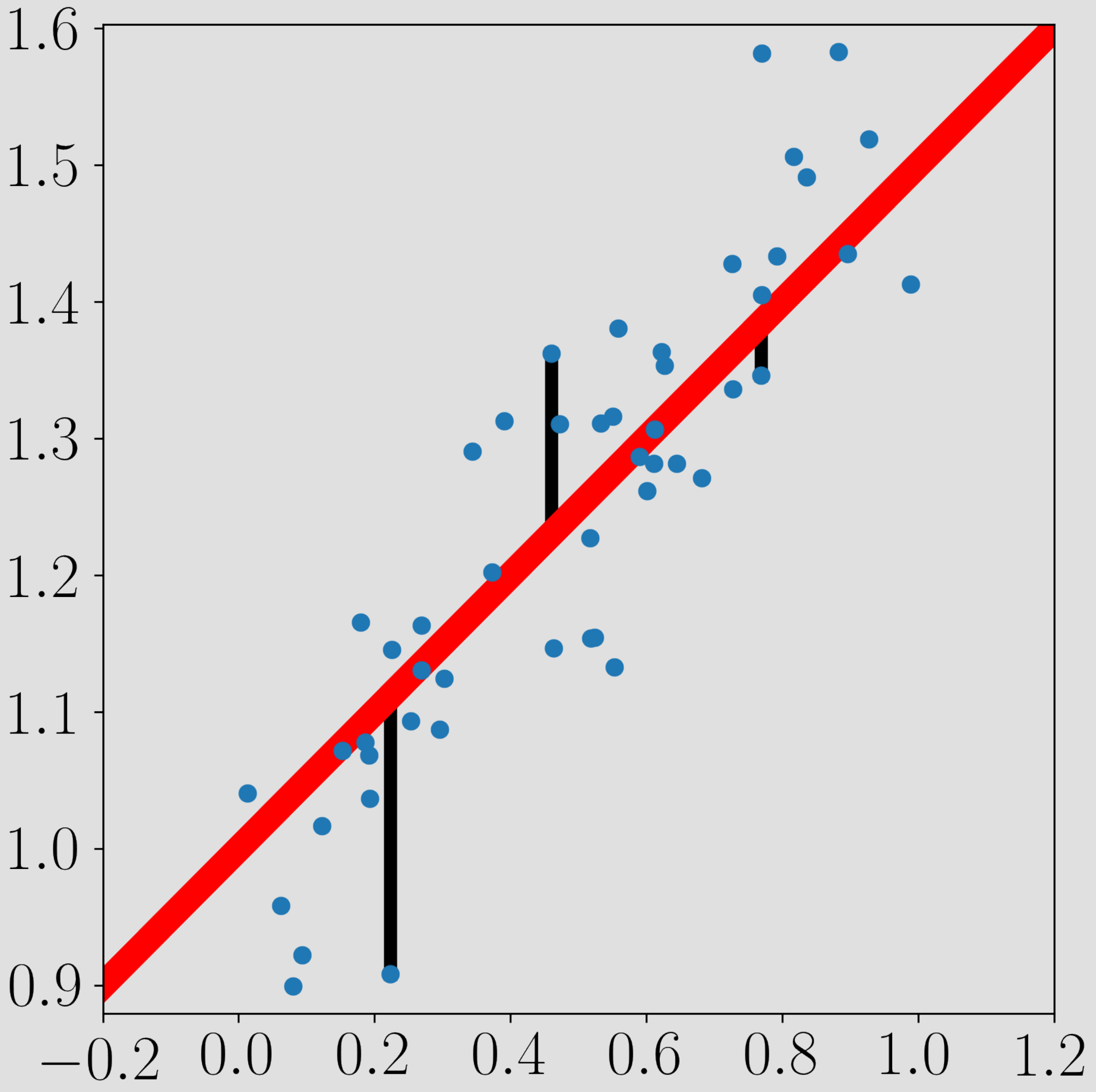
Linear regression



Linear regression

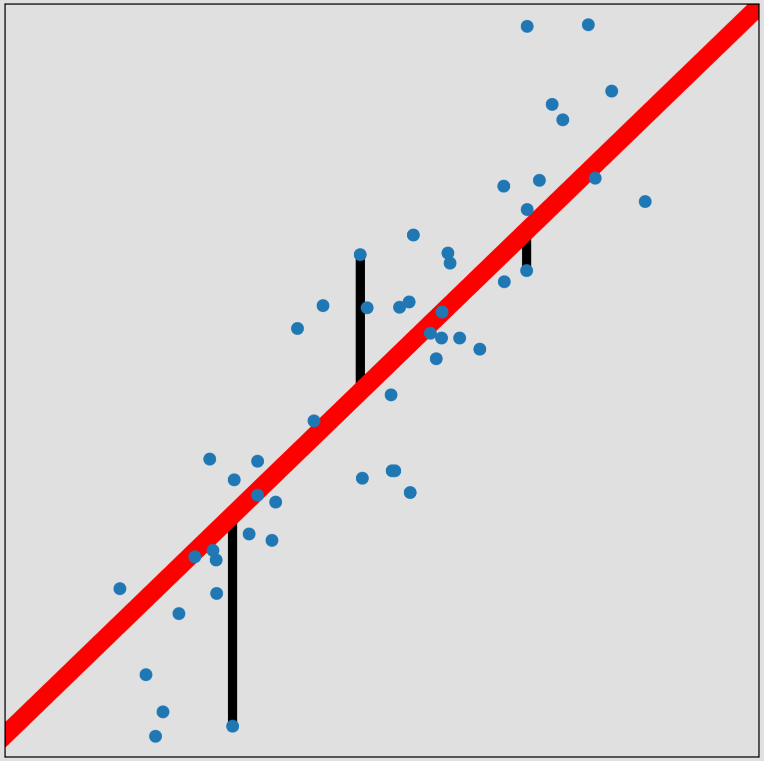


Linear regression



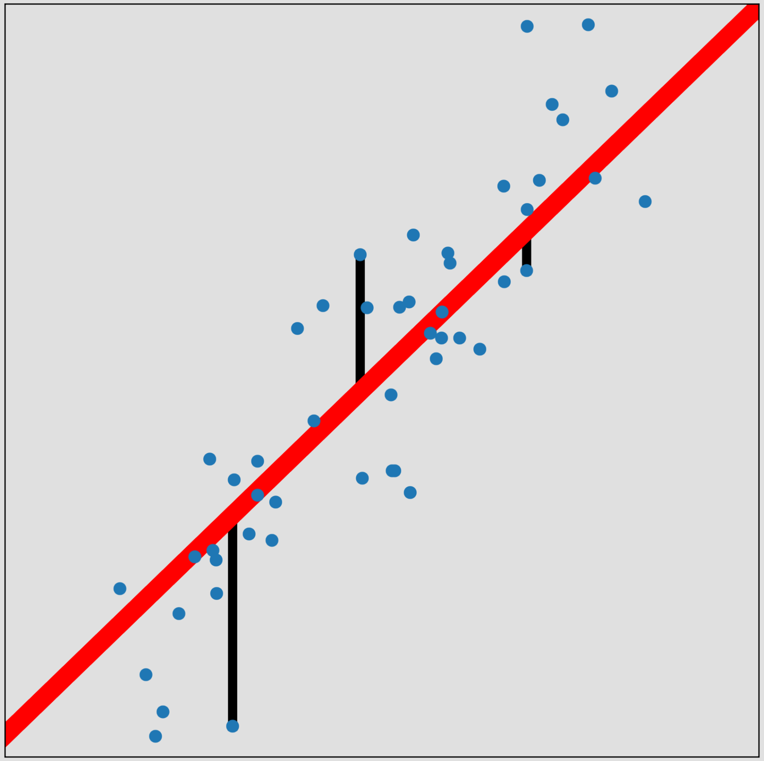
Least squares problem

$$L(w) = \sum_{i=1}^{N} (w^T x_i - y_i)^2 = \left\| w^T X - y \right\|^2 \to \min_w$$



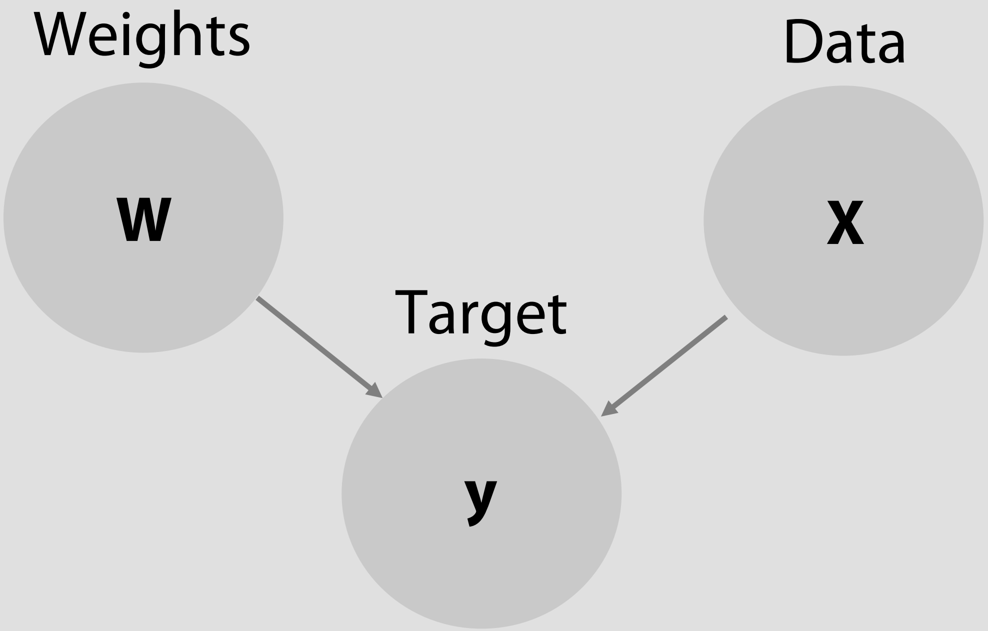
Least squares problem

$$L(w) = \sum_{i=1}^{N} (w^T x_i - y_i)^2 = \|w^T X - y\|^2 \to \min_w$$

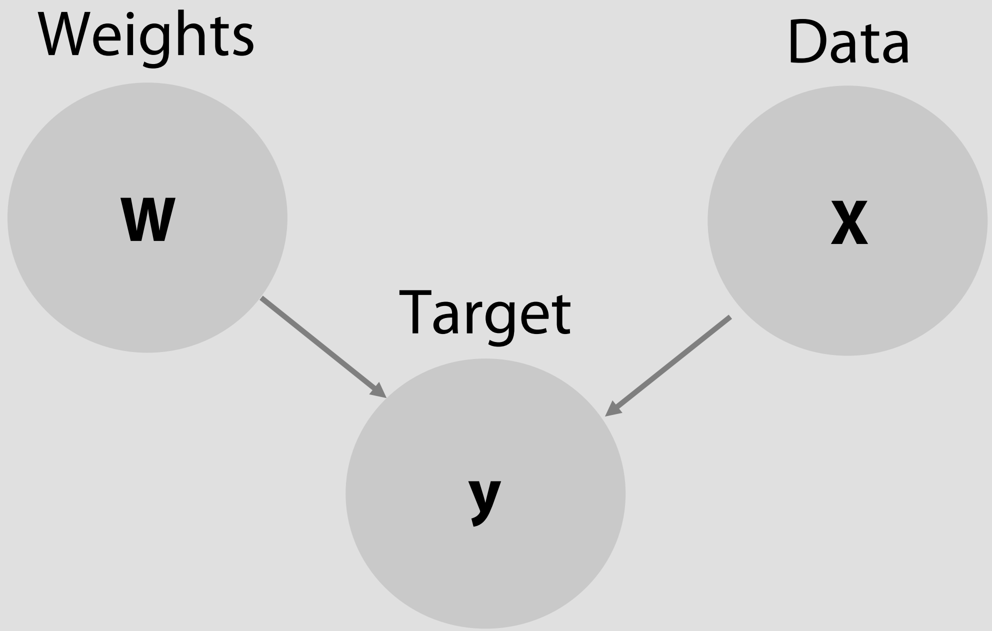


$$\widehat{w} = arg \min_w L(w)$$

Model

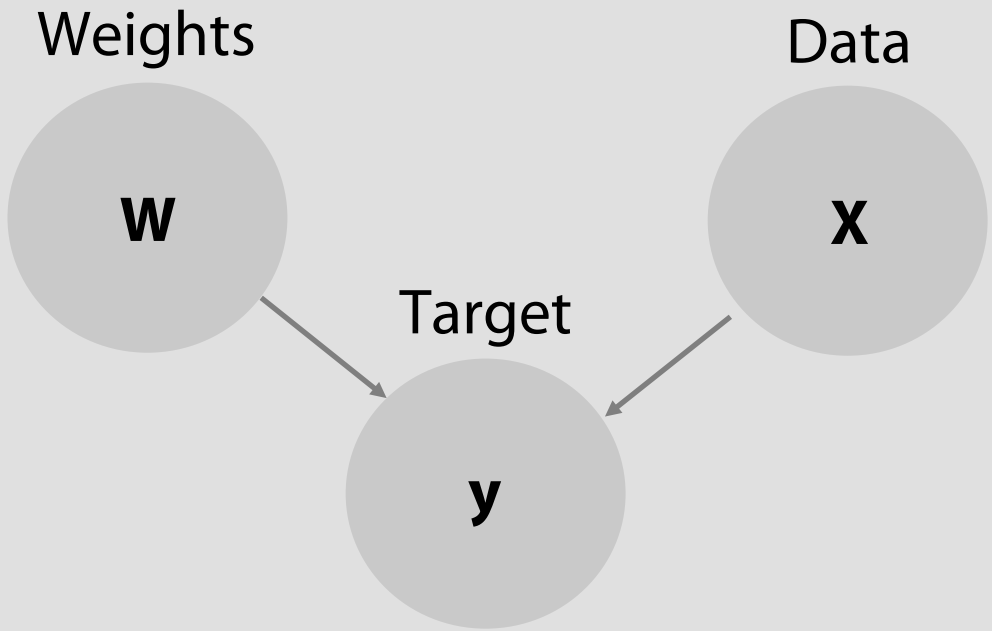


Model



$$P(w, y|X) = P(y|X, w)P(w)$$

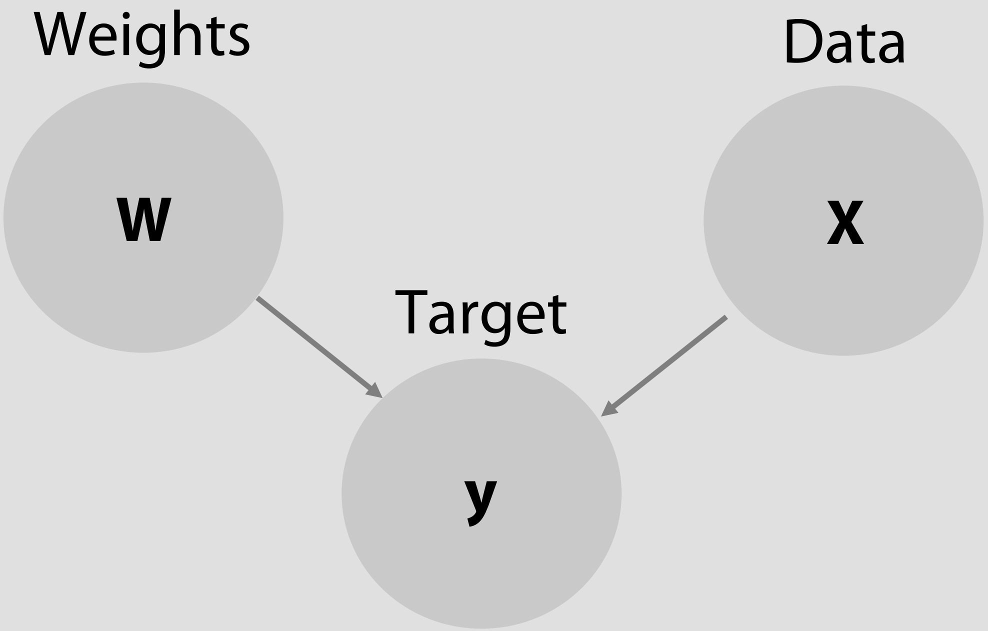
Model



$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

Model



$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ), 6 минут

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$P(w|y, X) = ?$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ)

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$P(w|y, X) = \frac{P(w, y|X)}{P(y|X)} \propto P(w, y|X)$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ)

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$P(w|y, X) = \frac{P(w, y|X)}{P(y|X)} \propto P(w, y|X)$$

$$P(w, y|X) \to \max_w$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ)

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$P(w|y, X) = \frac{P(w, y|X)}{P(y|X)} \propto P(w, y|X)$$

$$P(w, y|X) \to \max_w \iff \log P(w, y|X) \to \max_w$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ)

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$\log P(w, y|X) \to \max_w$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ)

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$\log P(w, y|X) \to \max_w$$

$$-\frac{1}{2\sigma^2}||w^T X - y||^2 - \frac{1}{2\gamma^2}||w||^2 \to \max_w$$

Training ТЕХНИЧЕСКИЙ СЛАЙД (НА ДОСКЕ)

$$P(w, y|X) = P(y|X, w)P(w)$$

$$P(y|w, X) = \mathcal{N}(y|w^T X, \sigma^2 I)$$

$$P(w) = \mathcal{N}(w|0, \gamma^2 I)$$

$$\log P(w, y|X) \to \max_w$$

$$-\frac{1}{2\sigma^2}||w^T X - y||^2 - \frac{1}{2\gamma^2}||w||^2 \to \max_w$$

$$||w^T X - y||^2 + C||w||^2 \to \min_w$$