Regression method	Simplest model: Sampling of one Gaussian variable x	Generalized linear regression (coloured Gaussian noise)	Generalized Linear regression (white Gaussian noise)
Least Squares	$\hat{\theta}_{LS} = \bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$	$\hat{\boldsymbol{\theta}}_{LS} = \left(\Phi^T\Phi\right)^{-1}\Phi^T \boldsymbol{y}$	$\hat{\boldsymbol{\theta}}_{LS} = \left(\Phi^T\Phi\right)^{-1}\Phi^T \boldsymbol{y}$
Ridge Regression	$\hat{\theta}_{RR} = \frac{N}{N+\lambda}\bar{x}, \quad \lambda_* = \frac{\sigma_\eta^2}{\theta_0^2}$	$\hat{\boldsymbol{\theta}}_{RR} = \left(\Phi^T\Phi + \lambda I\right)^{-1}\Phi^T \boldsymbol{y}$	$\hat{\boldsymbol{\theta}}_{RR} = \left(\Phi^T\Phi + \lambda \mathbf{I}\right)^{-1}\Phi^T \boldsymbol{y}$
Maximum Likelihood	• $p(x\|\theta) \to N(\theta, \sigma_\eta^2)$ $\hat{\theta}_{ML} = \bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0, \Sigma_\eta)$ $\hat{\boldsymbol{\theta}}_{ML} = \left(\Phi^T\Sigma_\eta^{-1}\Phi\right)^{-1}\Phi^T\Sigma_\eta^{-1}\boldsymbol{y}$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0, \sigma_\eta^2 I)$ $\hat{\boldsymbol{\theta}}_{ML} = \hat{\boldsymbol{\theta}}_{LS} = \left(\Phi^T\Phi\right)^{-1}\Phi^T \boldsymbol{y}$
Maximum a Posteriori	• $p(x\|\theta) \to N(\theta, \sigma_\eta^2)$ • Prior: $p(\theta) \to N(\theta_0, \sigma_\theta^2)$ $\hat{\theta}_{MAP} = \frac{N\sigma_\theta^2\bar{x} + \sigma_\eta^2\theta_0}{N\sigma_\theta^2 + \sigma_\eta^2}, \quad \bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0, \Sigma_\eta)$ • Prior: $p(\boldsymbol{\theta}) \to N(\boldsymbol{\theta}_0, \Sigma_\theta)$ $\hat{\boldsymbol{\theta}}_{MAP} = \boldsymbol{\theta}_0 + \left(\Sigma_\theta^{-1} + \Phi^T\Sigma_\eta^{-1}\Phi\right)^{-1}\Phi^T\Sigma_\eta^{-1}\left(\boldsymbol{y} - \Phi\boldsymbol{\theta}_0\right)$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0, \sigma_\eta^2 I)$ • Prior: $p(\boldsymbol{\theta}) \to N(\boldsymbol{\theta}_0, \sigma_\theta^2 I)$ $\hat{\boldsymbol{\theta}}_{MAP} = \boldsymbol{\theta}_0 + \frac{1}{\sigma_\eta^2}\left(\frac{1}{\sigma_\theta^2}I + \frac{1}{\sigma_\eta^2}\Phi^T\Phi\right)^{-1}\Phi^T\left(\boldsymbol{y} - \Phi\boldsymbol{\theta}_0\right)$
Bayesian Inference	• $p(x\|\theta) \to N(\theta, \sigma_\eta^2)$ • Prior: $p(\theta) \to N(\theta_0, \sigma_\theta^2)$ • Posterior: $p(\theta\|X) \to N(\theta_N, \sigma_N^2)$ $\theta_N = \frac{N\sigma_\theta^2\bar{x} + \sigma_\eta^2\theta_0}{N\sigma_\theta^2 + \sigma_\eta^2}, \sigma_N^2 = \frac{\sigma_\eta^2\sigma_\theta^2}{N\sigma_\theta^2 + \sigma_\eta^2}$ $p(x\|X) \to N(\theta_N, \sigma_x^2), \quad \sigma_x^2 = \sigma_\eta^2 + \frac{\sigma_\eta^2\sigma_N^2}{\sigma_N^2 + \sigma_\eta^2}$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0, \Sigma_\eta)$ • Prior: $p(\boldsymbol{\theta}) \to N(\boldsymbol{\theta}_0, \Sigma_\theta)$ • Posterior: $p(\boldsymbol{\theta} \mid \boldsymbol{y}) \to N(\boldsymbol{\theta} \mid \mu_{\theta\|y}, \Sigma_{\theta\|y})$ $\mu_{\theta\|y} = \boldsymbol{\theta}_0 + \left(\Sigma_\theta^{-1} + \Phi^T\Sigma_\eta^{-1}\Phi\right)^{-1}\Phi^T\Sigma_\eta^{-1}\left(\boldsymbol{y} - \Phi\boldsymbol{\theta}_0\right)$ $\Sigma_{\theta\|y} = \left(\Sigma_\theta^{-1} + \Phi^T\Sigma_\eta^{-1}\Phi\right)^{-1}$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0, \sigma_\eta^2 I)$ • Prior: $p(\boldsymbol{\theta}) \to N(\boldsymbol{\theta}_0, \sigma_\theta^2 I)$ • Posterior: $p(\boldsymbol{\theta} \mid \boldsymbol{y}) \to N(\boldsymbol{\theta} \mid \mu_{\theta\|y}, \Sigma_{\theta\|y})$ $\mu_{\theta\|y} = \boldsymbol{\theta}_0 + \frac{1}{\sigma_\eta^2}\left(\frac{1}{\sigma_\theta^2}I + \frac{1}{\sigma_\eta^2}\Phi^T\Phi\right)^{-1}\Phi^T\left(\boldsymbol{y} - \Phi\boldsymbol{\theta}_0\right)$ $\Sigma_{\theta\|y} = \left(\frac{1}{\sigma_\theta^2}I + \frac{1}{\sigma_\eta^2}\Phi^T\Phi\right)^{-1}$ $p(y \mid \boldsymbol{y}) \to N(y \mid \mu_y, \sigma_y^2)$ $\mu_y = \phi^T(x)\mu_{\theta\|y}, \sigma_y^2 = \sigma_\eta^2 + \sigma_\eta^2\sigma_\theta^2\phi^T(x)\left(\sigma_\eta^2 I + \sigma_\theta^2\Phi^T\Phi\right)^{-1}\phi(x)$

Regression method	Simplest model: Sampling of one Gaussian variable x	Generalized linear regression (white Gaussian noise)
Expectation maximization	• $p(x \mid \theta) \to N(\theta,\sigma_\eta^2)$ • Prior: $p(\theta) \to N(0,\sigma_\theta^2)$ • Posterior: $p(\theta \mid X) \to N(\mu_{\theta\mid x},\sigma^2_{\theta\mid x})$ • $\mu_{\theta\mid x} = \frac{N\sigma_\theta^2\bar{x}}{N\sigma_\theta^2+\sigma_\eta^2} = \frac{Nb\bar{x}}{Nb+a}, \sigma^2_{\theta\mid x} = \frac{\sigma_\eta^2\sigma_\theta^2}{N\sigma_\theta^2+\sigma_\eta^2} = \frac{1}{Nb+a}$ • Goal is to estimate: $a = 1/\sigma_\theta^2, b = 1/\sigma_\eta^2$ Initialize: $a^{(0)}, b^{(0)}$ While $\lvert a^{(j+1)} - a^{(j)} \rvert > \varepsilon, \lvert b^{(j+1)} - b^{(j)} \rvert > \varepsilon$ E-step: $\mu^{(j)}_{\theta\mid x} = \frac{Nb^{(j)}\bar{x}}{Nb^{(j)}+a^{(j)}}, \left(\sigma^{(j)}_{\theta\mid x}\right)^2 = \frac{1}{Nb^{(j)}+a^{(j)}}$ $A^{(j)} = \left(\mu^{(j)}_{\theta\mid x}\right)^2 + \left(\sigma^{(j)}_{\theta\mid x}\right)^2$ $B^{(j)} = \sum_{i=1}^{N}\left(x_i - \mu^{(j)}_{\theta\mid x}\right)^2 + N\left(\sigma^{(j)}_{\theta\mid x}\right)^2$ $Q(a,b,a^{(j)},b^{(j)}) = \frac{1}{2}\ln a + \frac{N}{2}\ln b - \frac{a}{2}A^{(j)} - \frac{b}{2}B^{(j)} + \text{constant}$ M-step: $\frac{\partial Q}{\partial a} = 0 \Rightarrow a^{(j+1)} = \frac{1}{A^{(j)}}$ $\frac{\partial Q}{\partial a} = 0 \Rightarrow b^{(j+1)} = \frac{N}{B^{(j)}}$	• Noise model: $\boldsymbol{\eta} = \boldsymbol{y} - \Phi\boldsymbol{\theta} \to N(0,\sigma_\eta^2 I)$ • Prior: $p(\theta) \to N(0,\sigma_\theta^2)$ • Posterior: $p(\boldsymbol{\theta} \mid \boldsymbol{y}) \to N(\boldsymbol{\mu}_{\theta\mid y},\Sigma_{\theta\mid y})$ $\mu_{\theta\mid y} = b\Sigma_{\theta\mid y}\Phi^T\boldsymbol{y}, \Sigma_{\theta\mid y} = \left(aI + b\Phi^T\Phi\right)^{-1}$ • Goal is to estimate: $a = 1/\sigma_\theta^2, b = 1/\sigma_\eta^2$ Initialize: $a^{(0)}, b^{(0)}$ While $\lvert a^{(j+1)} - a^{(j)} \rvert > \varepsilon, \lvert b^{(j+1)} - b^{(j)} \rvert > \varepsilon$ E-step: $\mu^{(j)}_{\theta\mid y} = b^{(j)}\Sigma^{(j)}_{\theta\mid y}\Phi^T\boldsymbol{y}, \Sigma^{(j)}_{\theta\mid y} = \left(a^{(j)}I + b^{(j)}\Phi^T\Phi\right)^{-1}$ $A^{(j)} = \left\lVert \boldsymbol{\mu}^{(j)}_{\theta\mid y} \right\rVert^2 + \text{trace}\left\{\Sigma^{(j)}_{\theta\mid y}\right\}$ $B^{(j)} = \left\lVert \boldsymbol{y} - \Phi\boldsymbol{\mu}^{(j)}_{\theta\mid y} \right\rVert^2 + \text{trace}\left\{\Phi\Sigma^{(j)}_{\theta\mid y}\Phi^T\right\}$ $Q(a,b,a^{(j)},b^{(j)}) = \frac{K}{2}\ln a + \frac{N}{2}\ln b - \frac{a}{2}A^{(j)} - \frac{b}{2}B^{(j)} + \text{constant}$ M-step: $\frac{\partial Q}{\partial a} = 0 \Rightarrow a^{(j+1)} = \frac{K}{A^{(j)}}$ $\frac{\partial Q}{\partial a} = 0 \Rightarrow b^{(j+1)} = \frac{N}{B^{(j)}}$