Lecture THIRTEEN

$$y_t = \beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \cdots + \beta_{n-1}(t-(n-1))_+ + \varepsilon_t$$

$$\beta_0, \beta_1 \overset{iid}{\sim} \underbrace{N(0, C)}_{\text{or } \mathrm{Unif}(-C,C)}, \quad \beta_2, \ldots, \beta_{n-1} \overset{iid}{\sim} N(0, \tau^2)$$

$\log \tau \sim \mathrm{Unif}(-C, C)$, $\log \sigma \sim \mathrm{Unif}(-C,C)$

$$\underbrace{\beta}_{\beta_0, \beta_1, \ldots, \beta_{n-1}}, \sigma, \tau$$

What is the posterior?

$$\underbrace{\underset{n\times 1}{\beta} \,\Big|\, \text{data}, \sigma, \tau}_{\text{posterior of } \beta \text{ when } \sigma, \tau \text{ are fixed}} \sim N\left(\left(\frac{X^T X}{\sigma^2} + Q^{-1}\right)^{-1} \frac{X^T y}{\sigma^2}, \left(\frac{X^T X}{\sigma^2} + Q^{-1}\right)^{-1}\right)$$

Here

$$Q = \begin{bmatrix} C & & & & 0 \\ & C & & & \\ & & \tau^2 & & \\ & & & \ddots & \\ 0 & & & & \tau^2 \end{bmatrix}$$

$$Q^{-1} = \begin{bmatrix} 1/C & & & & 0 \\ & 1/C & & & \\ & & 1/\tau^2 & & \\ & & & \ddots & \\ 0 & & & & 1/\tau^2 \end{bmatrix} \approx \begin{bmatrix} 0 & & & & 0 \\ & 0 & & & \\ & & 1/\tau^2 & & \\ & & & \ddots & \\ 0 & & & & 1/\tau^2 \end{bmatrix} = \frac{1}{\tau^2} J$$

$$J = \begin{bmatrix} 0 & & & & 0 \\ & 0 & & & \\ & & 1 & & \\ & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}$$

$$\beta \,\Big|\, \substack{\text{data} \\ \tau,\sigma} \sim N\left(\left(\frac{X^TX}{\sigma^2} + \frac{J}{\tau^2}\right)^{-1} \frac{X^Ty}{\sigma^2}, \left(\frac{X^TX}{\sigma^2} + \frac{J}{\tau^2}\right)^{-1} \right)$$

$$= N\left(\left(X^TX + \frac{\sigma^2}{\tau^2} J\right)^{-1} X^Ty, \; \sigma^2 \left(X^TX + \frac{\sigma^2}{\tau^2} J\right)^{-1} \right)$$

Ridge Regression:

$$\underbrace{\|y - X\beta\|^2 + \lambda \sum_{j=2}^{n-1} \beta_j^2}$$

$$\hat{\beta}_{\text{ridge}}(\lambda) = \left(X^TX + \lambda J\right)^{-1} X^Ty$$

$$= \text{posterior mean of } \beta \,\Big|\, \substack{\text{data} \\ \sigma,\tau}$$

provided $\lambda = \frac{\sigma^2}{\tau^2}$.

$\boxed{f_{\beta|\sigma,\tau,\text{data}}}$

$\updownarrow$

$\boxed{f(\sigma,\tau) \quad \sigma,\tau \,|\, \text{data}}$

$$\propto \frac{\sigma^{-n-1}\tau^{-1}}{\sqrt{\det Q}} \sqrt{\det\left(\frac{X^TX}{\sigma^2} + Q^{-1}\right)^{-1}} \exp\left(-\frac{y^Ty}{2\sigma^2}\right) \exp\left(\frac{y^TX}{2\sigma^2}\left(\frac{X^TX}{\sigma^2} + Q^{-1}\right)^{-1} \frac{X^Ty}{\sigma^2}\right)$$

$$Q = \operatorname{diag}(C, C, \tau^2, \ldots, \tau^2)$$

$$\det Q = C^2 (\tau^2)^{n-2} \propto (\tau^2)^{n-2}$$

$$Q^{-1} \approx J/\tau^2$$

Evaluate the posterior of (τ,σ) over a grid of values of (τ,σ).

① Generate posterior samples of τ,σ

② Generate β given τ, σ.

Most case: ⓐ $f_{\tau,\sigma|data}(\tau,\sigma)$ prefers τ which are not too large.

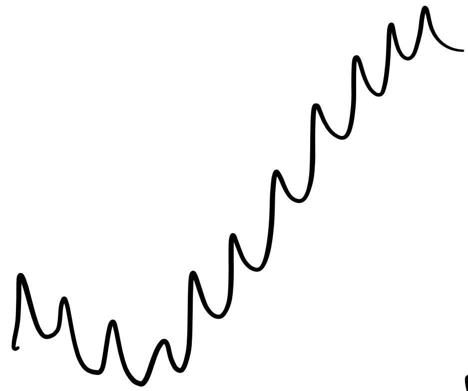
(AVOID OVERFITTING)

$\tau = 0.5$

$\sigma = 2$

$f_{\tau,\sigma|data}(0.5, 2) = 17$

$f_{\tau,\sigma|data}(0.05, 2) = 25$



ⓑ $f_{\tau,\sigma|data}(\tau,\sigma)$ prefers τ which are not too small

(AVOID UNDERFITTING)

$$\boxed{f_{\tau,\sigma|data}(\tau,\sigma)} \propto \boxed{f_{data|\tau,\sigma}(data)}\; \underbrace{f_{\tau,\sigma}(\tau,\sigma)}_{\boxed{\frac{1}{\tau\sigma}}}$$

$\log \tau \sim Unif[C,C]$

$\log \sigma \sim Unif[C,C]$

Uninformative

$$\boxed{f_{data|\tau,\sigma}(data)}$$

Likelihood: $f_{data|\beta,\sigma}(data)$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{\|y - X\beta\|^2}{2\sigma^2}\right)$$

$$\boxed{f_{data|\tau,\sigma}(data)} = \int f_{data|\beta,\sigma,\tau}(data)\, f_{\beta|\tau,\sigma}(\beta)\, d\beta$$

Integrated Likelihood (Marginal Likelihood)

$$\underbrace{f_{data|\tau,\sigma}(data)}_{\text{Integrated Likelihood}}$$

$$\underbrace{f_{data|\beta,\sigma}(data)}_{\text{Original likelihood}}$$

Will be large for values of β which lead to overfitting

$$f_{data|\tau,\sigma}(data) = \int \underbrace{f_{data|\beta,\sigma}(data)} \underbrace{\boxed{f_{\beta|\tau}(\beta)}} d\beta$$

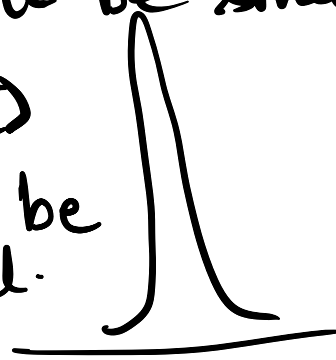
$$\frac{1}{\sqrt{2\pi}\,\tau} \exp\left(-\frac{\beta_j^2}{2\tau^2}\right)$$

① τ large: $N(0,\tau^2)$

usually will be small.

② τ small: $N(0,\tau^2)$

↳ $f_{data|\tau,\sigma}(data)$ will be small.



Slightly Different Prior

τ, σ

Reparametrize $\tau = \sigma \times \gamma$

Change the prior to

$\log\tau, \log\sigma \overset{iid}{\sim} \text{Unif}(-C, C)$

$\Big\downarrow$ CHANGE

$$\underbrace{\log\gamma, \log\sigma \overset{iid}{\sim} \text{Unif}(-C, C)}$$

allows tractable integration of σ in $f_{\sigma,\gamma|\text{data}}(\sigma, \gamma)$.

Ridge: $\lambda = \dfrac{\sigma^2}{\tau^2}$, $\quad \tau = \sigma \times \gamma.$

$\boxed{\lambda = \dfrac{1}{\gamma^2}}$ or $\boxed{\gamma = \dfrac{1}{\sqrt{\lambda}}}$

Posterior: (β, σ, γ)

$$\beta \,\Big|\, \begin{matrix}\text{data} \\ \sigma, \gamma\end{matrix} \sim N\left(\left(\frac{X^TX}{\sigma^2} + Q^{-1}\right)^{-1} \frac{X^Ty}{\sigma^2}, \left(\frac{X^TX}{\sigma^2} + Q^{-1}\right)^{-1}\right)$$

$$Q = \begin{bmatrix} C & & & 0 \\ & C & & \\ & & \sigma^2\gamma^2 & \\ & & & \ddots \\ 0 & & & \sigma^2\gamma^2 \end{bmatrix}$$

$f_{\sigma,\gamma|\text{data}}$:

$$f_{\gamma|\text{data}}(\gamma) = \frac{\gamma^{-n+1}\left|X^TX+\gamma^{-2}J\right|^{-1/2}}{\left(y^Ty - y^TX\left(X^TX+\gamma^{-2}J\right)^{-1}X^Ty\right)^{\frac{n}{2}-1}}$$

$$\boxed{\sigma \,|\, \gamma, \text{data}}$$

$$\frac{1}{\sigma^2}\,\Big|\,\text{data}, \gamma \sim \text{Gamma}\left(\frac{n}{2}-1, \frac{y^Ty - y^TX(X^TX+\gamma^{-2}J)^{-1}X^Ty}{2}\right)$$

① First take a grid for γ & compute posterior. Samples for γ.

② ~~For~~ For each γ sample, generate σ.

③ Given γ & σ, generate β.

VARIANCE MODELS (SPECTRAL ANALYSIS)

ALL THE MODELS we studied so far are examples of <u>mean</u> models.

<u>e.g</u>:

$$y_t = \boxed{\beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \cdots + \beta_{n-1}(t-(n-1))_+} + \varepsilon_t$$

$$\boxed{\sum\left[y_t - (\quad)\right]^2 + \lambda \sum_j \beta_j^2}$$

or $\lambda \sum_j |\beta_j|$

$$y_t = \mu_t + \varepsilon_t$$

$$\sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_{j=2}^{n-1} \Big(\underbrace{(\mu_j - \mu_{j-1}) - (\mu_{j-1} - \mu_{j-2})} \Big)^2$$

$$\sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_{t=2}^{n-1} \left| (\mu_t - \mu_{t-1}) - (\mu_{t-1} - \mu_{t-2}) \right|$$

Data: $y_1, \ldots, y_n$

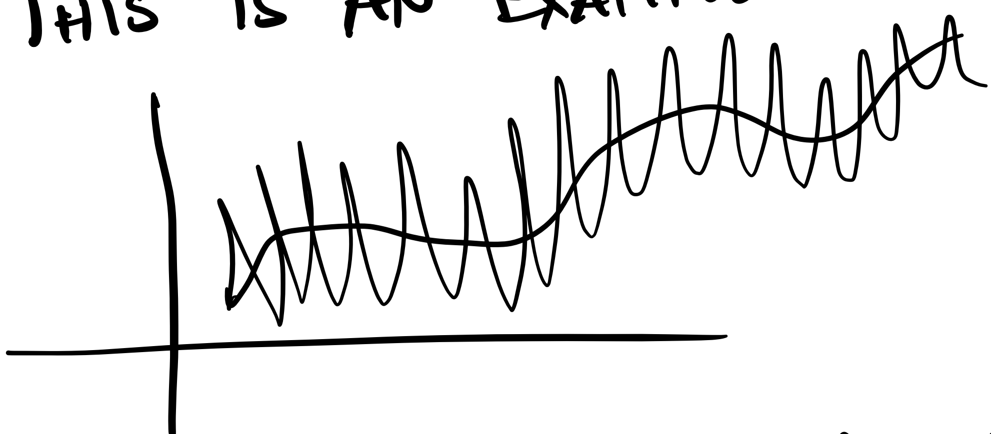
Model: $y_t \overset{\text{independent}}{\sim} N(\underline{\mu_t}, \sigma^2)$

<u>Goal</u>: To estimate the means $\mu_1 \ldots \mu_n$

<u>Estimate</u>: Assume smoothness.

$$\Sigma(y_t - \mu_t)^2 + \lambda \sum_{t=2}^{n-1} \left[(\mu_t - \mu_{t-1}) - (\mu_{t-1} - \mu_{t-2}) \right]^2$$

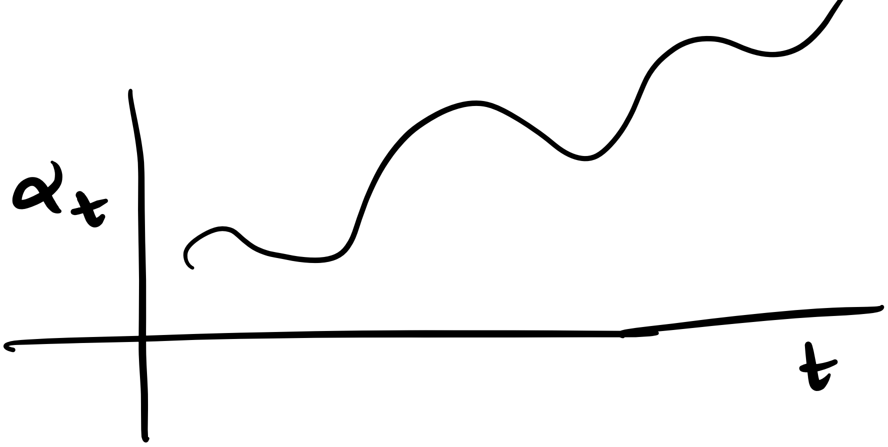
THIS IS AN EXAMPLE OF A MEAN MODEL



In Contrast, Variance Model example:

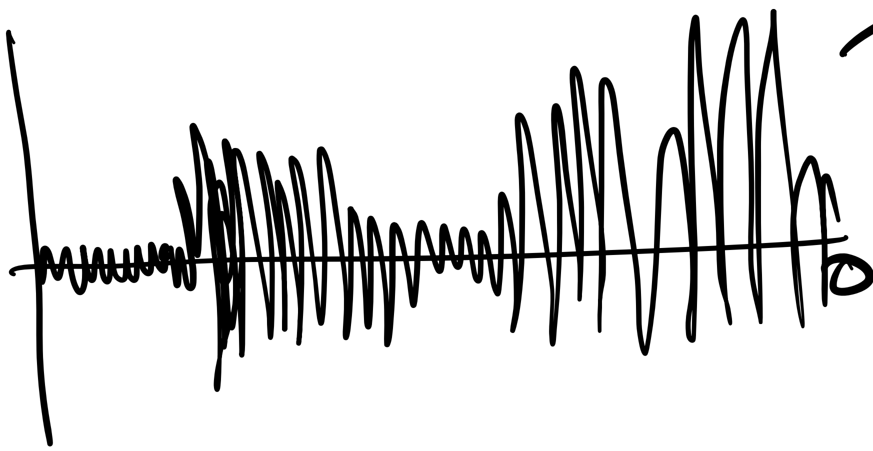
$y_1, \ldots y_n$

$\left[y_t \overset{\text{independent}}{\sim} N(0, \sigma_t^2) \right., \quad t = 1 \ldots n$



$$\sigma_t = \exp(\alpha_t)$$

$$y_t \overset{\text{indepedent}}{\sim} N(0, \sigma_t^2)$$



Variance model.