ESERCIZIO 10

The task is to analyze a dataset where observations d_i depend linearly on predictors $x_i = (x_{i1}, \dots, x_{ip})$ through the model:

$$d_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i$$

The noise ε_i follows a Gaussian distribution $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

LASSO REGRESSION

We assume that each parameter β_j is independently distributed according to a Laplace prior:

$$p(\beta_j) = \frac{1}{2b} \exp\left(-\frac{|\beta_j|}{b}\right)$$

The negative log of this prior is $-\log p(\beta_j) = \frac{|\beta_j|}{b} + \text{const}$

The likelihood is:

$$p(\vec{d} \mid \vec{\beta}, \vec{x}) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2\right)$$

$$-\log p(\vec{d} \mid \vec{\beta}, \vec{x}) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2 + \text{const}$$

The posterior is proportional to the product of the likelihood and the prior:

$$p(\vec{\beta} \mid \vec{d}, \vec{x}) \propto p(\vec{d} \mid \vec{\beta}, \vec{x}) \cdot \prod_{j=1}^{p} p(\beta_j)$$

The MAP estimate minimizes the negative log-posterior

$$-\log p(\vec{\beta} \mid \vec{d}, \vec{x}) \propto -\log\left[p(\vec{d} \mid \vec{\beta}, \vec{x}) \cdot \prod_{j=1}^{p} p(\beta_j)\right]$$

$$\propto -\log\left[p(\vec{d} \mid \vec{\beta}, \vec{x})\right] + \left(-\log\left[\prod_{j=1}^{p} p(\beta_j)\right]\right) =$$

$$\propto \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2 + \frac{1}{b} \sum_{j=1}^{p} |\beta_j| + \text{const}$$

So the MAP estimate is:

$$\hat{\vec{\beta}} = \arg\min_{\vec{\beta}} \left(\sum_i^m \left(d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}\right)^2 + \frac{2\sigma^2}{b} \sum_{k=0}^p |\beta_k| \right) =$$

$$= \boxed{\arg\min \left(RSS + \frac{2\sigma^2}{b} \sum_{k=0}^p |\beta_k| \right)}$$

This is the formula of the LASSO REGRESSION

RIDGE REGRESSION

Now we assume a Gaussian prior:

$$p(\beta_j) = \frac{1}{\sqrt{2\pi c}} \exp\left(-\frac{\beta_i^2}{2c}\right)$$

$$-\log p(\beta_j) = \frac{\beta_j^2}{2c} + \text{const}$$

We use the same likelihood of the first part

$$-\log P(\vec{\beta} \mid \vec{d}, \vec{x}) \propto \frac{1}{2\sigma^2} \sum_{i=1}^m \left(d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}\right)^2 + \frac{1}{2c} \sum_{j=1}^p \beta_j^2 + \text{const}$$

So the MAP estimate is:

$$\hat{\vec{\beta}} = \arg\min_{\vec{\beta}} \left(\sum_i^m \left(d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}\right)^2 + \frac{\sigma^2}{c} \sum_{k=0}^p \beta_k^2 \right) =$$

$$= \boxed{\arg\min_{\vec{\beta}} \left(RSS + \frac{\sigma^2}{c} \sum_{k=0}^p \beta_k^2 \right)}$$

This formula is called RIDGE REGRESSION

ESERCIZIO 11

We have a likelihood:

$$L(\theta, x) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} \exp\left[-\frac{1}{2}(x - A\bar{x})^T C^{-1}(x - A\bar{x})\right]$$

We want to do a marginalization den A assuming a uniform prior $p(A) = 1$

The posterior of θ after marginalizing A is:

$$P(\theta|x) \propto \int_{-\infty}^{\infty} L(\theta, x)\, p(A)\, dA$$

$$\propto \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(x - A\bar{x})^T C^{-1}(x - A\bar{x})\right] dA$$

$$(x - A\bar{x})^T C^{-1}(x - A\bar{x}) = x^T C^{-1} x - 2A x^T C^{-1}\bar{x} + A^2 \bar{x}^T C^{-1}\bar{x} =$$
$$= A^2 \bar{x}^T C^{-1}\bar{x} - 2A(x^T C^{-1}\bar{x}) + x^T C^{-1} x$$

I rewrite in the form $\gamma A^2 - 2\delta A - \eta$, where $\gamma = \bar{x}^T C^{-1}\bar{x}$, $\delta = x^T C^{-1}\bar{x}$ and $\eta = x^T C^{-1} x$

To facilitate marginalization, we complete the square in A

$$\gamma\left(A^2 - 2\frac{\delta}{\gamma}A + \frac{\delta^2}{\gamma^2}\right) + \eta - \frac{\delta^2}{\gamma} = \gamma\left(A - \frac{\delta}{\gamma}\right)^2 + \eta - \frac{\delta^2}{\gamma}$$

The likelihood becomes

$$L(\theta, x) \propto \exp\left(-\frac{1}{2}\left[\gamma\left(A - \frac{\delta}{\gamma}\right)^2 + \eta - \frac{\delta^2}{\gamma}\right]\right)$$
$$\propto \exp\left(-\frac{\eta}{2}\right)\exp\left(\frac{\delta^2}{2\gamma}\right)\exp\left(-\frac{\gamma}{2}\left(A - \frac{\delta}{\gamma}\right)^2\right)$$

The terms involving A is a Gaussian integral:

$$\int_{-\infty}^{\infty} \exp\left(-\frac{\gamma}{2}\left(A - \frac{\delta}{\gamma}\right)^2\right) dA = \sqrt{\frac{2\pi}{\gamma}}$$

So the posterior becomes:

$$P(\theta \mid x) \propto \exp\left(-\frac{M}{2}\right) \exp\left(\frac{\delta^2}{2\gamma}\right) \sqrt{\frac{2\pi}{\gamma}}$$

$$P(\theta \mid x) \propto \frac{1}{\sqrt{\bar{x}^T C^{-1} \bar{x}}} \exp\left(-\frac{1}{2}\left[x^T C^{-1} x - \frac{(x^T C^{-1} \bar{x})^2}{\bar{x}^T C^{-1} \bar{x}}\right]\right)$$

The marginalization over A effectively projects the data x onto the defined direction by $\bar{x}$, weighted by the precision matrix C^{-1}.

ESERCIZIO 12

The data d_i at time t_i is modelled as:

$$d_i = f(t_i) + n_i$$

where $f(t_i) = B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)$, $\quad n_i \sim N(0, \sigma^2)$

STEP 1

The likelihood of the data given the parameters is:

$$L(D|\omega, B_1, B_2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(d_i - f(t_i))^2\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{N}(d_i^2 + f(t_i)^2 - 2d_i f(t_i))}_{:Q}\right)$$

We know that $f(t_i) = B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)$, so:

$$f(t_i)^2 = B_1^2 \cos^2(\omega t_i) + B_2^2 \sin^2(\omega t_i) + 2B_1 B_2 \cos(\omega t_i)\sin(\omega t_i)$$

$$2d_i f(t_i) = 2d_i B_1 \cos(\omega t_i) + 2d_i B_2 \sin(\omega t_i)$$

$$\therefore Q = \sum_{i=1}^{N} d_i^2 - 2B_1 \sum_{i=1}^{N} d_i \cos(\omega t_i) - 2B_2 \sum_{i=1}^{N} d_i \sin(\omega t_i) +$$

$$+ B_1^2 \sum_{i=1}^{N} \cos^2(\omega t_i) + B_2^2 \sum_{i=1}^{N} \sin^2(\omega t_i)$$

$$R(\omega) = \sum_{i=1}^{N} d_i \cos(\omega t_i)$$

$$I(\omega) = \sum_{i=1}^{N} d_i \sin(\omega t_i)$$

$$c = \sum_{i=1}^{N} \cos^2(\omega t_i)$$

$$S = \sum_{i=1}^{N} \sin^2(\omega t_i)$$

$$Q = \sum_{i=1}^{N} d_i^2 - 2B_1 R(\omega) - 2B_2 I(\omega) + B_1^2 C + B_2^2 S$$

STEP 2 : MARGINALIZING OVER B_1 AND B_2

The likelihood becomes:

$$L(D|\omega) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{Q}{2\sigma^2}\right) dB_1 \, dB_2$$

Separate Q into terms dependent on B_1 and B_2:

$$Q = \text{constant} + c\left(B_1 - \frac{R(\omega)}{c}\right)^2 - \frac{R(\omega)^2}{c} + s\left(B_2 - \frac{I(\omega)}{s}\right)^2 - \frac{I(\omega)^2}{s}$$

The terms $\left(B_1 - \frac{R(\omega)}{c}\right)^2$ and $\left(B_2 - \frac{I(\omega)}{s}\right)^2$ are Gaussian integrals. Their contributions are:

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{c}{2\sigma^2}\left(B_1 - \frac{R(\omega)}{c}\right)^2\right) dB_1 \propto \sqrt{\frac{2\pi\sigma^2}{c}}$$

and similarly for B_2

So:

$$P(\omega|D) \propto \frac{1}{\sqrt{cs}} \exp\left(\frac{R(\omega)^2}{2\sigma^2 c} + \frac{I(\omega)^2}{2\sigma^2 s}\right)$$

STEP 3 : PERIODOGRAM

After the marginalization the posterior is $P(\omega|D) \propto \frac{1}{\sqrt{cs}} \exp\left(\frac{R(\omega)^2}{2\sigma^2 c} + \frac{I(\omega)^2}{2\sigma^2 s}\right)$

Because ωt_i is uniformly distributed over $[0, 2\pi]$, we can do the approximations:

$$\sum_{i=1}^{N} \cos^2(\omega t_i) \approx \frac{N}{2} \qquad \sum_{i=1}^{N} \sin^2(\omega t_i) \approx \frac{N}{2}$$

$$\Rightarrow \quad c \approx \frac{N}{2} \qquad s \approx \frac{N}{2}$$

$$\Rightarrow P(\omega|D) \propto \frac{1}{N} \exp\left(\frac{R(\omega)^2}{N\sigma^2} + \frac{I(\omega)^2}{N\sigma^2}\right)$$

We know that $R(\omega) = \sum_{i=1}^{N} d_i \cos(\omega t_i)$ and $I(\omega) = \sum_{i=1}^{N} d_i \sin(\omega t_i)$

The power spectrum:

$$P(\omega) = R(\omega)^2 + I(\omega)^2$$

So the posterior is $P(\omega|D) \propto \exp\left(\frac{P(\omega)}{N\sigma^2}\right)$

We can write also $P(\omega) = \left(\sum_{i=1}^{N} d_i \cos(\omega t_i)\right)^2 + \left(\sum_{i=1}^{N} d_i \sin(\omega t_i)\right)^2$

Thus correspon to the squared magnitude of the Fourier transform:

$$P(\omega) = \left|\sum_{i=1}^{N} d_i e^{-i\omega t_i}\right|^2$$

$$\Rightarrow C(\omega) = \frac{1}{N} \left|\sum_{i=1}^{N} d_i e^{-i\omega t_i}\right|^2$$

The MAP estimate of ω maximizes the posterior $P(\omega|D)$. Since the exponential function monotonically increases with its argument, maximizing the posterior is equivalent to maximizing the periodogram:

$$\hat{\omega} = \arg\max_{\omega} C(\omega)$$

STEP 4: LEAST SQUARES FITTING

The least squares fitting corresponds to minimizing $\chi^2 = Q/\sigma^2$. This ensures consistency with the Gaussian likelihood under the assumptions:

1. Independent Gaussian noise in the data
2. A linear dependence of the model on ω over small intervals

If $\omega t_i \approx 0$ the sinusoidal model can be approximated as:

$$f(t_i) \approx B_1 + B_2 \omega t_i$$

resulting in a linear relationship.