Home Work 1, due Sunday, Oct 20, 11:59 pm

Consider polynomial regression where we regress over powers of x, so we look at model M_n of the form

$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_n x^n + \sigma^2 \mathcal{N}(0, I), n = 0, 1, 2, \ldots, K-1$$

Given a data set $\mathcal{D} = \{(y_i, x_i)\}_{i=0,1,2,\ldots,N-1}$. We want to compute

$$\underset{n=0,1,\ldots,K-1}{\operatorname{argmax}} \; p(M_n|\mathcal{D})$$

By Bayes' formula

$$p(M_n|\mathcal{D}) = \frac{p(\mathcal{D}|M_n)p(M_n)}{p(\mathcal{D})}$$

We set the prior $p(M_n) = 1/K$, $n = 0, 1, 2, \ldots, K-1$

Thus

$$p(M_n|\mathcal{D}) = \frac{p(\mathcal{D}|M_n)}{Kp(\mathcal{D})}$$

so

$$\underset{n}{\operatorname{argmax}} \; p(M_n|\mathcal{D}) = \underset{n}{\operatorname{argmax}} \; p(\mathcal{D}|M_n)$$

$p(\mathcal{D}|M_n)$ is known as the *Model Evidence.* We can express it as a marginal distribution where we marginalize over the parameters of the model

$$p(\mathcal{D}|M_n) = \int p(\mathcal{D}|w, M_n)p(w|M_n)dw$$

a) Verify this formula

Let X_n denote the matrix

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N-1} & x_{N-1}^2 & \dots & x_{N-1}^n \end{pmatrix}$$

Then

p(w|Mn) p(w|Mn)

X should be also ind of Mn

$$p(\mathcal{D}|w, M_n) = \frac{p(y, X, w, M_n)}{p(w, M_n)} = p(y|X, w, M_n)p(X, w|M_n)$$

X and w are independent so $p(X, w|M_n) = P(X)p(w|M_n)$ where $p(w|M_n)$ is a prior on the $n+1$-dimensional parameter vector w.

$$p(y|X, w, M_n) = \mathcal{N}(\cancel{w^T X_n}, \sigma^2 I_N)$$

Xn*w

Thus

$$p(\mathcal{D}|M_n) = p(X) \int \mathcal{N}(y; \cancel{w^T X_n}, \sigma^2 I_{N-1})p(w|M_n)dw$$

We put a $\mathcal{N}(0, \Sigma_p)$ prior on w, where Σ_p is an $n \times n$ positive definite covariance matrix.

b) Show that

$$\frac{1}{\sigma^2}(y-X_n w)^T(y-X_n w)+w^T\Sigma_p^{-1}w = (w-\mu)^T A^{-1}(w-\mu)+\frac{1}{\sigma^2}\left(y^T y - \frac{1}{\sigma^2}y^T X_n A^T X_n^T y\right)$$

where

$$A = \left(\frac{1}{\sigma^2}X_n^T X_n + \Sigma_p^{-1}\right)^{-1}$$

and

$$\mu = \left(\frac{1}{\sigma^2}X_n^T X_n + \Sigma_p^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}X_n^T y\right) = A\left(\frac{1}{\sigma^2}X_n^T y\right)$$

We have

$$\mathcal{N}(y; X_n w, \sigma^2 I_N) = \frac{1}{2\pi^{N/2}} \frac{1}{(\det \sigma^2 I_N)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(y - X_n w)^T (y - X_n w)\right)$$

$$\mathcal{N}(w; 0, \Sigma_p) = \frac{1}{2\pi^{(n+1)/2}} \frac{1}{(\det \Sigma_p)^{1/2}} \exp\left(-\frac{1}{2}(w^T \Sigma_p^{-1} w)\right)$$

and

$$\mathcal{N}(w; A\left(\frac{1}{\sigma^2} X_n^T y\right), A)$$

$$= \frac{1}{2\pi^{(n+1)/2}} \frac{1}{(\det A)^{1/2}} \exp\left(-\frac{1}{2}\left(w - A\left(\frac{1}{\sigma^2}X_n^T y\right)\right)^T A^{-1} \left(w - A\left(\frac{1}{\sigma^2}X_n^T y\right)\right)\right)$$

c) Show

$$\mathcal{N}(y; X_n w, \sigma^2 I_N)\mathcal{N}(w; 0, \Sigma_p) = \frac{1}{2\pi^{N/2}} \frac{1}{(\det \sigma^2 I_N)^{1/2}} \frac{1}{2\pi^{(n+1)/2}} \frac{1}{(\det \Sigma_p)^{1/2}}$$
$$(2\pi)^{(n+1)/2} (\det A)^{1/2} \exp\left(-\frac{1}{2} y^T \left(\frac{1}{\sigma^2} I_N - \frac{1}{\sigma^4} X_n^{\cancel{T}} A^{T} X_n^{T}\right) y\right)$$
$$\mathcal{N}(w; A\left(\frac{1}{\sigma^2} X_n^T y\right), A)$$

d) Show

$$\int \mathcal{N}(y; X_n w, \sigma^2 I_n)\mathcal{N}(w; 0, \Sigma_p) = \frac{\sqrt{\det A}}{(2\pi)^{N/2}\sigma^N \sqrt{\det \Sigma_p}} \exp\left(-\frac{1}{2} y^T \left(\frac{1}{\sigma^2} I_N - \frac{1}{\sigma^4} X_n^{\cancel{T}} A^{T} X_n^{T}\right) y\right)$$

e) Use the data set `polynomial_reg.csv` and compute R^2 for polynomial regressions of degrees $1, 2, 3, 4, 5, 6, 7, 8$. Use $\sigma^2 = 1$ and $\Sigma_p = 0.85 I_{n+1}$ and the formulas derived in the previous sections to compute $\operatorname{argmax}_n p(M_n|\mathcal{D})$ and run the regression for that degree. You may find it helpful to use the package

`sklearn.preprocessing PolynomialFeatures`