PDF-XChange Click to buy NOW! www.tracker-software.com
PDF-XChange Click to buy NOW! www.tracker-software.com

Machine Learning, Homework 2 (Learning algorithm- Gradient Descent method)

Due date: 2021/04/12

For a given periodic function $f(x)=f(x+T)$, we can use the Fourier series to approximate it as follows

$$f(x)=\frac{A_0}{2}+\sum_{n=1}^{N}A_n\cos\left(\frac{2\pi n}{T}x+\phi_n\right)=a_0+\sum_{n=1}^{N}\left(a_n\cos\frac{2\pi n}{T}x+b_n\sin\frac{2\pi n}{T}x\right) \tag{1}$$

where the corresponding Fourier series coefficients are

$$a_n=\frac{2}{T}\int_{x_0}^{x_0+T}f(x)\cdot\cos\frac{2\pi n}{T}x\,dx \tag{2a}$$

$$b_n=\frac{2}{T}\int_{x_0}^{x_0+T}f(x)\cdot\sin\frac{2\pi n}{T}x\,dx\,. \tag{2b}$$

Herein, we would like to obtain the Fourier series coefficients by gradient descent method with square error cost function. The general used learning algorithm is gradient descent method is

$$W\leftarrow W+\left(-\alpha\frac{\partial E(\cdot)}{\partial W}\right)$$

where $E(k)$ is the error cost function and W is the adjustable parameters. The error cost function is $E(k)=\frac{1}{2}\left(y(k)-\hat{y}(k)\right)^2$ and $E=\frac{1}{2}\sum_{k=1}^{M}\left(y(k)-\hat{y}(k)\right)^2$ for pattern learning and batch learning, respectively. M denotes the data number. We can achieve it by the following steps.

(i) At first, select a periodic (non-sinusoidal) function, $f(x)=f(x+T)$

(ii) Obtain the input/output data (x, $f(x)$), obtain the training pattern by randomly choose or uniformly choose.

(iii) Define the approximation model $\hat{f}(x(k))=\hat{a}_0+\sum_{n=1}^{N}\left(\hat{a}_n\cos\frac{2\pi n}{T}x(k)+\hat{b}_n\sin\frac{2\pi n}{T}x(k)\right)$ and error cost function $E(k)=\frac{1}{2}\left(f(x(k))-\hat{f}(x(k))\right)^2$ and $E=\frac{1}{2}\sum_{k=1}^{M}\left(f(x(k))-\hat{f}(x(k))\right)^2$.

Please answer the following questions.

(a) According the gradient descent method, please derive the update laws for $\hat{a}_n$ and $\hat{b}_n$, n=0,1,….

(b) Implement the learning algorithm to find the corresponding $\hat{a}_n$ and $\hat{b}_n$. Are the values of $\hat{a}_n$ and $\hat{b}_n$ equal to equation (2a) and (2b)? Please give the corresponding learning parameters and MSE.

(c) Give a detailed discussion for *order N* vs. *mean square error*, $\text{MSE}=\frac{1}{M}\sum_{k=1}^{M}\left(y(k)-\hat{y}(k)\right)^2$.

PDF-XChange Product
Click to buy NOW!
www.tracker-software.com
PDF-XChange Product
Click to buy NOW!
www.tracker-software.com

(d) Give a detailed discussion for *learning rate* vs. *mean square error*.

(e) Give a brief discussion for pattern learning and batch learning with the same learning rate.

(f) Consider the gradient descent method with momentum

$$W \leftarrow W + V_t$$

$$V_t \leftarrow \beta V_{t-1} + \left(-\alpha \frac{\partial E}{\partial W} \right)$$

Repeat part (b) and compare these two methods. Please give your observation.

(g) In actual system, the measure signals usually having sensor noise. Please add noise with different signal noise ratio (SNR), and repeat part (b) to obtain the analysis (SNR vs. MSE).

(h) As above, it can be viewed as $f(x)$ is approximated by linear combination of sinusoidal basis. Would we use polynomial basis to treat it? Why?