CSC 412 Machine Learning and Knowledge Discovery

Exercise I

1. **Linear Regression**

After we trained a regression model, we got a parameter vector:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} -1 \\ -1.5 \\ +3 \end{bmatrix}$$

Now we have a test sample:

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{y}$
3	2	1

(a) What is the sample's feature vector $\boldsymbol{x}$?

(b) What is $\hat{\boldsymbol{y}}$?

(c) What is the MSE of this test set?

(d) What is the $\boldsymbol{\theta}^{(\text{next step})}$, if we regard $\hat{\boldsymbol{\theta}}$ as the current step, and this sample is the single instance picked by Stochastic Gradient Descent? Let the learning rate be 0.1 (Hint: $\nabla = \boldsymbol{X}^\intercal(\boldsymbol{X\theta} - \boldsymbol{y})$)

(e) List all the new features, if we transform the data to fit a Fourth-Degree Polynomial Regression.

Answer:

(a)

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

(b)

$$\hat{\boldsymbol{y}} = \boldsymbol{\theta}^\intercal \boldsymbol{x} = \begin{bmatrix} -1 & -1.5 & +3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = (-1)\cdot 1 + (-1.5)\cdot 3 + 3 \cdot 2 = -1 - 4.5 + 6 = 0.5$$

(c)

$$\text{MSE} = \frac{1}{1}\sum_{i=1}^{1}(\hat{y}^{(i)} - y^{(i)})^2 = (0.5-1)^2 = 0.25$$

(d)

$$\begin{aligned}
\boldsymbol{\theta}^{(\text{next step})} &:= \boldsymbol{\theta} - \eta \cdot \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \\
&= \boldsymbol{\theta} - \eta \cdot \Big(\boldsymbol{X}^\intercal (\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}) \Big) \\
&= \begin{bmatrix} -1 \\ -1.5 \\ +3 \end{bmatrix} - 0.1 \cdot \{ \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \cdot (\begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -1.5 \\ +3 \end{bmatrix} - \begin{bmatrix} 1 \end{bmatrix}) \} \\
&= \begin{bmatrix} -1 \\ -1.5 \\ +3 \end{bmatrix} - 0.1 \cdot \{ \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} (\begin{bmatrix} 0.5 \end{bmatrix} - \begin{bmatrix} 1 \end{bmatrix}) \} \\
&= \begin{bmatrix} -1 \\ -1.5 \\ +3 \end{bmatrix} - 0.1 \cdot \begin{bmatrix} -0.5 \\ -1.5 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} -1 \\ -1.5 \\ +3 \end{bmatrix} - \begin{bmatrix} -0.05 \\ -0.15 \\ -0.1 \end{bmatrix} \\
&= \begin{bmatrix} -0.95 \\ -1.35 \\ +3.1 \end{bmatrix}
\end{aligned}$$

(e)

$$\begin{array}{lllll}
\boldsymbol{x_1}^4 & \boldsymbol{x_2}^4 & \boldsymbol{x_1}^3\boldsymbol{x_2} & \boldsymbol{x_1}\boldsymbol{x_2}^3 & \boldsymbol{x_1}^2\boldsymbol{x_2}^2 \\
\boldsymbol{x_1}^3 & \boldsymbol{x_2}^3 & \boldsymbol{x_1}^2\boldsymbol{x_2} & \boldsymbol{x_1}\boldsymbol{x_2}^2 & \\
\boldsymbol{x_1}^2 & \boldsymbol{x_2}^2 & \boldsymbol{x_1}\boldsymbol{x_2} & & \\
\boldsymbol{x_1} & \boldsymbol{x_2} & & & \\
\boldsymbol{x_0} & & & &
\end{array}$$

2. **Logistic Regression Cost Function**

Let the Logistic Regression cost function be

$$J(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(\hat{p}^{(i)}\right) + (1-y^{(i)})\log\left(1-\hat{p}^{(i)}\right)\right]$$

Then what is the cost of this test result (table below)?

$\hat{\boldsymbol{p}}$	$\boldsymbol{y}$
0.2	0
0.4	1
0.6	0
0.8	1

Answer:

$$\begin{aligned}
J = -\frac{1}{4} \cdot \big[\, & 0 \cdot \log(0.2) + (1-0)\cdot\log(1-0.2) \\
& + 1 \cdot \log(0.4) + (1-1)\cdot\log(1-0.4) \\
& + 0 \cdot \log(0.6) + (1-0)\cdot\log(1-0.6) \\
& + 1 \cdot \log(0.8) + (1-1)\cdot\log(1-0.8)\,\big] \\
= -\frac{1}{4} & \cdot (\log 0.8 + \log 0.4 + \log 0.4 + \log 0.8) \\
= -\frac{1}{2} & \log 0.8 - \frac{1}{2}\log 0.4
\end{aligned}$$