Lehrstuhl für STEUERUNGS- UND REGELUNGSTECHNIK Technische Universität München Prof. Dongheui Lee	MACHINE LEARNING IN ROBOTICS Exercises 1: Linear Regression

Exercise 1

Given the dataset shown in table 1 and illustrated in figure 1, we want to predict the output value for $x = 1$.

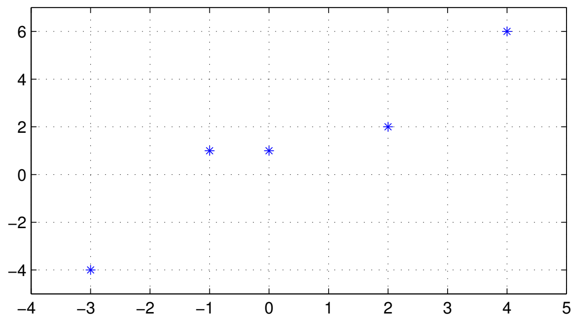

Figure 1: Training dataset

input x	-3	-1	0	2	4
output y	-4	1	1	2	6

Table 1: Data

a) Let's assume $f(x) = w_1x + w_2x^2$ is a regression model with unknown parameter vector $\boldsymbol{w} = [w_1 \quad w_2]^T$. Find $\boldsymbol{w}$ which fits the data best in the sense of the Euclidean norm.

b) Predict the output value of the system for $x = 1$.

Solution Exercise 1

a)

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} -3 & 9 \\ -1 & 1 \\ 0 & 0 \\ 2 & 4 \\ 4 & 16 \end{bmatrix}, \ \boldsymbol{y} = \begin{bmatrix} -4 \\ 1 \\ 1 \\ 2 \\ 6 \end{bmatrix}$$

$$\boldsymbol{w}^* = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T y}$$

$$\boldsymbol{X^T X} = \begin{bmatrix} -3 & -1 & 0 & 2 & 4 \\ 9 & 1 & 0 & 4 & 16 \end{bmatrix} \begin{bmatrix} -3 & 9 \\ -1 & 1 \\ 0 & 0 \\ 2 & 4 \\ 4 & 16 \end{bmatrix} = \begin{bmatrix} 30 & 44 \\ 44 & 354 \end{bmatrix}$$

$$(\boldsymbol{X^T X})^{-1} = \frac{1}{10620 - 1936} \begin{bmatrix} 354 & -44 \\ -44 & 30 \end{bmatrix} = \frac{1}{8684} \begin{bmatrix} 354 & -44 \\ -44 & 30 \end{bmatrix}$$

$$\boldsymbol{X^T y} = \begin{bmatrix} -3 & -1 & 0 & 2 & 4 \\ 9 & 1 & 0 & 4 & 16 \end{bmatrix} \begin{bmatrix} -4 \\ 1 \\ 1 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 39 \\ 69 \end{bmatrix}$$

$$(\boldsymbol{X^T X})^{-1} \boldsymbol{X^T y} = \frac{1}{8684} \begin{bmatrix} 354 & -44 \\ -44 & 30 \end{bmatrix} \begin{bmatrix} 39 \\ 69 \end{bmatrix} = \frac{1}{8684} \begin{bmatrix} 13806 - 3036 \\ -1716 + 2070 \end{bmatrix}$$

$$= \frac{1}{8684} \begin{bmatrix} 10770 \\ 354 \end{bmatrix} = \begin{bmatrix} 1.2402 \\ 0.0408 \end{bmatrix}$$

b) $x = 1 \Longrightarrow y = 1.2402 * 1 + 0.0408 * 1^2 = 1.281$.

Exercise 2

Consider the following sum-of-squares error function:

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}\left(f(x^{(i)}, \boldsymbol{w}) - y^{(i)}\right)^2$$

in which the function $f(x^{(i)}, \boldsymbol{w})$ is the polynomial:

$$f(x^{(i)}, \boldsymbol{w}) = w_0 + w_1\left(x^{(i)}\right) + w_2\left(x^{(i)}\right)^2 + \ldots + w_m\left(x^{(i)}\right)^m = \sum_{j=0}^{m} w_j\left(x^{(i)}\right)^j \ .$$

Show that the coefficients $\boldsymbol{w} = \{w_k\}$ that minimize this error function are given by the solution to the following set of linear equations

$$\sum_{j=0}^{m} A_{kj} w_j = Y_k$$

where

$$A_{kj} = \sum_{i=1}^{n}(x^{(i)})^{k+j}, \qquad Y_k = \sum_{i=1}^{n}(x^{(i)})^k y^{(i)} \ .$$

Here a suffix i or j denotes the index of a component, whereas $(x)^i$ denotes x to the power of i.

Solution Exercise 2

Substituting $f(x^{(i)}, \boldsymbol{w})$ into $E(\boldsymbol{w})$ we obtain

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{j=0}^{m} w_j\left(x^{(i)}\right)^j - y^{(i)}\right)^2$$

The optimal coefficients $\boldsymbol{w} = \{w_k\}$ are then computed by setting to zero the derivatives $E(\boldsymbol{w})$ with respect to w_k

$$\frac{\partial E(\boldsymbol{w})}{\partial w_k} = \frac{2}{2}\sum_{i=1}^{n}\left(\sum_{j=0}^{m} w_j\left(x^{(i)}\right)^j - y^{(i)}\right)\left(x^{(i)}\right)^k = \sum_{i=1}^{n}\left(\sum_{j=0}^{m} w_j\left(x^{(i)}\right)^j - y^{(i)}\right)\left(x^{(i)}\right)^k = 0, \quad \forall k$$

The previous equation can be rewritten as

$$\begin{aligned}\sum_{i=1}^{n}\left(\sum_{j=0}^{m} w_j\left(x^{(i)}\right)^j - y^{(i)}\right)\left(x^{(i)}\right)^k &= \sum_{i=1}^{n}\left(\sum_{j=0}^{m} w_j\left(x^{(i)}\right)^j\left(x^{(i)}\right)^k\right) - \sum_{i=1}^{n} y^{(i)}\left(x^{(i)}\right)^k \\ &= \sum_{j=0}^{m}\sum_{i=1}^{n} w_j\left(x^{(i)}\right)^{k+j} - \sum_{i=1}^{n} y^{(i)}\left(x^{(i)}\right)^k = 0\end{aligned}$$

Choosing

$$A_{kj} = \sum_{i=1}^{n}(x^{(i)})^{k+j}, \qquad Y_k = \sum_{i=1}^{n}(x^{(i)})^k y^{(i)}$$

we have

$$\sum_{j=0}^{m} A_{kj} w_j = Y_k$$

Exercise 3

The kinematic of a differential-drive mobile robot like that in figure 2 is described in the discrete-time by the set of equations

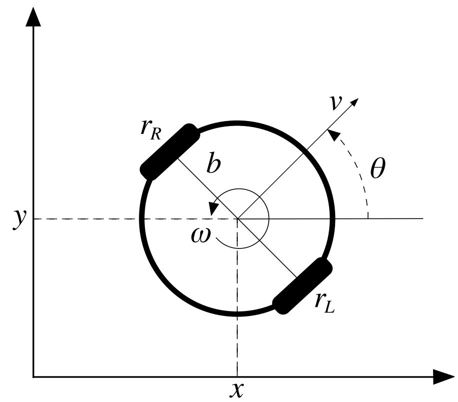

Figure 2: Top-view sketch of a differential-drive mobile robot with relevant variables.

$$\begin{cases} x^{(t+1)} = x^{(t)} + v^{(t)} cos(\theta^{(t)} + \omega^{(t)} \frac{\Delta T}{2}) \Delta T \\ y^{(t+1)} = y^{(t)} + v^{(t)} sin(\theta^{(t)} + \omega^{(t)} \frac{\Delta T}{2}) \Delta T \\ \theta^{(t+1)} = \theta^{(t)} + \omega^{(t)} \Delta T \end{cases}$$

where ΔT is the sample time. The relation between the linear v and angular ω velocities of the robot and the velocity of the wheels (ω_R and ω_L) is

$$\begin{bmatrix} v \\ \omega \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \omega_R \\ \omega_L \end{bmatrix} = \boldsymbol{W} \begin{bmatrix} \omega_R \\ \omega_L \end{bmatrix}$$

Given m motion trajectories $T_r = \left[\left\{ x_1^{(t)}, y_1^{(t)}, \theta_1^{(t)}, \omega_{R,1}^{(t)}, \omega_{L,1}^{(t)} \right\}_{t=0}^{n}, \ldots, \left\{ x_m^{(t)}, y_m^{(t)}, \theta_m^{(t)}, \omega_{R,1}^{(t)}, \omega_{L,1}^{(t)} \right\}_{t=0}^{n} \right]$, estimate the unknown parameters $\boldsymbol{W}$ using least square regression. (Hint: $[w_{11}, w_{12}]$ and $[w_{21}, w_{22}]$ can be separately estimated.)

Solution Exercise 3

Let's rewrite the expression of $\theta^{(t+1)}$ to underline the dependence on the unknown parameters $[w_{21}, w_{22}]$. For $t = 0$ it holds that

$$\theta^{(1)} = \theta^{(0)} + w_{21} \Delta T \omega_R^{(0)} + w_{22} \Delta T \omega_L^{(0)} \ ,$$

for the final instant n it holds that

$$\theta^{(n)} = \theta^{(0)} + w_{21} \Delta T \sum_{t=0}^{n-1} \omega_R^{(t)} + w_{22} \Delta T \sum_{t=0}^{n-1} \omega_L^{(t)} \ ,$$

that can be written in a compact form by choosing $\boldsymbol{X}_\theta = \Delta T \left[\sum_{t=0}^{n-1} \omega_R^{(t)} \ \sum_{t=0}^{n-1} \omega_L^{(t)} \right]$

$$\theta^{(n)} - \theta^{(0)} = \boldsymbol{X}_\theta \begin{bmatrix} w_{21} \\ w_{22} \end{bmatrix}$$

The m given trajectories can be stacked in the form

$$\begin{bmatrix} \theta_1^{(n)} - \theta_1^{(0)} \\ \vdots \\ \theta_m^{(n)} - \theta_m^{(0)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{\theta,1} \\ \vdots \\ \boldsymbol{X}_{\theta,m} \end{bmatrix} \begin{bmatrix} w_{21} \\ w_{22} \end{bmatrix} = \bar{\boldsymbol{X}}_\theta \begin{bmatrix} w_{21} \\ w_{22} \end{bmatrix}$$

and the optimal values for $[w_{21}, w_{22}]$ estimated as

$$\begin{bmatrix} w_{21}^* \\ w_{22}^* \end{bmatrix} = \left(\bar{\boldsymbol{X}}_\theta^T \bar{\boldsymbol{X}}_\theta \right)^{-1} \bar{\boldsymbol{X}}_\theta^T \begin{bmatrix} \theta_1^{(n)} - \theta_1^{(0)} \\ \vdots \\ \theta_m^{(n)} - \theta_m^{(0)} \end{bmatrix} .$$

Following a similar reasoning it is possible to estimate $[w_{11}, w_{12}]$. Let's rewrite the expression of $x^{(t+1)}$ and $y^{(t+1)}$ for $t = 0$. To easy the notation, we substitute $\alpha^{(i)} = \theta^{(i)} + \omega^{(i)} \frac{\Delta T}{2}$.

$$\begin{cases} x^{(1)} = x^{(0)} + w_{11} \Delta T \omega_R^{(0)} cos(\alpha^{(0)}) + w_{21} \Delta T \omega_L^{(0)} cos(\alpha^{(0)}) \\ y^{(1)} = y^{(0)} + w_{11} \Delta T \omega_R^{(0)} sin(\alpha^{(0)}) + w_{21} \Delta T \omega_L^{(0)} sin(\alpha^{(0)}) \end{cases}$$

For the final instant n it holds that

$$\begin{cases} x^{(n)} - x^{(0)} = w_{11} \Delta T \displaystyle\sum_{t=0}^{n-1} \omega_R^{(t)} cos(\alpha^{(t)}) + w_{21} \Delta T \sum_{t=0}^{n-1} \omega_L^{(t)} cos(\alpha^{(t)}) \\ y^{(n)} - y^{(0)} = w_{11} \Delta T \displaystyle\sum_{t=0}^{n-1} \omega_R^{(t)} sin(\alpha^{(t)}) + w_{21} \Delta T \sum_{t=0}^{n-1} \omega_L^{(t)} sin(\alpha^{(t)}) \end{cases}$$

that can be written in the compact form

$$\begin{bmatrix} x^{(n)} - x^{(0)} \\ y^{(n)} - y^{(0)} \end{bmatrix} = \boldsymbol{X}_{xy} \begin{bmatrix} w_{11} \\ w_{12} \end{bmatrix} .$$

by choosing

$$\boldsymbol{X}_{xy} = \Delta T \begin{bmatrix} \sum_{t=0}^{n-1} \omega_R^{(t)} cos(\alpha^{(t)}) & \sum_{t=0}^{n-1} \omega_L^{(t)} cos(\alpha^{(t)}) \\ \sum_{t=0}^{n-1} \omega_R^{(t)} sin(\alpha^{(t)}) & \sum_{t=0}^{n-1} \omega_L^{(t)} sin(\alpha^{(t)}) \end{bmatrix} .$$

The m given trajectories can be stacked in the form

$$\begin{bmatrix} x_1^{(n)} - x_1^{(0)} \\ y_1^{(n)} - y_1^{(0)} \\ \vdots \\ x_m^{(n)} - x_m^{(0)} \\ y_m^{(n)} - y_m^{(0)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{xy,1} \\ \vdots \\ \boldsymbol{X}_{xy,m} \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \end{bmatrix} = \bar{\boldsymbol{X}}_{xy} \begin{bmatrix} w_{11} \\ w_{12} \end{bmatrix}$$

and the optimal values for $[w_{11}, w_{12}]$ estimated as

$$\begin{bmatrix} w_{11}^* \\ w_{12}^* \end{bmatrix} = \left(\bar{\boldsymbol{X}}_{xy}^T \bar{\boldsymbol{X}}_{xy}\right)^{-1} \bar{\boldsymbol{X}}_{xy}^T \begin{bmatrix} x_1^{(n)} - x_1^{(0)} \\ y_1^{(n)} - y_1^{(0)} \\ \vdots \\ x_m^{(n)} - x_m^{(0)} \\ y_m^{(n)} - y_m^{(0)} \end{bmatrix}.$$

Exercise 4

Given the dataset shown in table 2, we want to predict the output values for $x_1 = 7, x_2 = 4$. We assume a linear regression model.

input x_1	3	2	1	3
input x_2	5	2	2	3
output y_1	7	4	2	6
output y_2	10	5	9	6

Table 2: Data

a) Let's assume as a regression model:

$$f_1(\boldsymbol{x}) = w_0 x_1 + w_1 x_2$$
$$f_2(\boldsymbol{x}) = w_0^{'} x_1 + w_1^{'} x_2$$

with unknown parameters

$$\boldsymbol{w} = \begin{bmatrix} w_0 & w_0^{'} \\ w_1 & w_1^{'} \end{bmatrix}.$$

Find the best $\boldsymbol{w}$ using the normal equation.

b) Predict the output value of the system for $x_1 = 7, x_2 = 4$.

Solution

a) $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}$. Using the normal equation it is easy to compute the optimal parameters:

$$\boldsymbol{W}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{1}$$

where

$$\boldsymbol{X} = \begin{bmatrix} 3 & 5 \\ 2 & 2 \\ 1 & 2 \\ 3 & 3 \end{bmatrix}$$

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \left(\begin{bmatrix} 3 & 2 & 1 & 3 \\ 5 & 2 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 2 & 2 \\ 1 & 2 \\ 3 & 3 \end{bmatrix} \right)^{-1} = \left(\begin{bmatrix} 23 & 30 \\ 30 & 42 \end{bmatrix} \right)^{-1} = \frac{1}{66} \begin{bmatrix} 42 & -30 \\ -30 & 23 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6364 & -0.4545 \\ -0.4545 & 0.3485 \end{bmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 3 & 2 & 1 & 3 \\ 5 & 2 & 2 & 3 \end{bmatrix} \begin{bmatrix} 7 & 10 \\ 4 & 5 \\ 2 & 9 \\ 6 & 6 \end{bmatrix} = \begin{bmatrix} 49 & 67 \\ 65 & 96 \end{bmatrix}$$

$$\boldsymbol{W}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 0.6364 & -0.4545 \\ -0.4545 & 0.3485 \end{bmatrix} \begin{bmatrix} 49 & 67 \\ 65 & 96 \end{bmatrix} = \begin{bmatrix} 1.6364 & -1.0 \\ 0.3788 & 3 \end{bmatrix}$$

b) $x_1 = 7, x_2 = 4 \Longrightarrow \boldsymbol{y} = \boldsymbol{x}^T\boldsymbol{W} \Longrightarrow y_1 = 1.6364 * 7 + 0.3788 * 4 = 12.97, y_2 = -7 + 12 = 5.$