ECE C247 Homework 01

Chih-En Lin

1. (a)

(i) Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$\Rightarrow AA^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix}$$

$$= I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Rightarrow \begin{cases} a^2+b^2 = 1 \\ c^2+d^2 = 1 \\ ac+bd = 0 \end{cases}$$

Assume $a = \frac{1}{2}$, $d = \frac{-1}{2}$

$$\Rightarrow \begin{cases} b = \frac{\sqrt{3}}{2} \\ c = \frac{\sqrt{3}}{2} \end{cases} \Rightarrow A = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{bmatrix}$$

$$Ax = \lambda x$$

$$\Rightarrow (A - \lambda I)x = 0$$

$$\det(A - \lambda I) = 0$$

$$\Rightarrow \det\left(\begin{bmatrix} \frac{1}{2} - \lambda & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} - \lambda \end{bmatrix}\right) = 0$$

$$\Rightarrow \left(\frac{1}{2} - \lambda\right)\left(\frac{-1}{2} - \lambda\right) - \frac{3}{4} = 0$$

$$\Rightarrow \lambda^2 - 1 = 0$$

$$\Rightarrow \lambda = \pm 1$$

eigenvalues: 1, -1

① eigenvector with $\lambda_1 = 1$:

$(A - I)x = 0$

$$\Rightarrow \begin{bmatrix} \frac{1}{2}-1 & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2}-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} \frac{-1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-3}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$\Rightarrow x_1 = \sqrt{3}\, x_2$

$\Rightarrow$ Let $x_2 = 1 \Rightarrow x_1 = \sqrt{3}$

normalize

$$\Rightarrow x = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}, \quad \lambda_1 = 1$$

② eigenvector with $\lambda_2 = -1$

$(A + I)x = 0$

$$\Rightarrow \begin{bmatrix} \frac{1}{2}+1 & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2}+1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} \frac{3}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$\Rightarrow x_2 = -\sqrt{3}\, x_1$

$\Rightarrow$ Let $x_1 = 1 \Rightarrow x_2 = -\sqrt{3}$

normalize

$$\Rightarrow x = \begin{bmatrix} \frac{1}{2} \\ \frac{-\sqrt{3}}{2} \end{bmatrix}, \quad \lambda_2 = -1$$

We can notice that eigenvalues $|\lambda| = 1$,
eigenvectors are orthogonal. #

(ii) $AA^T = I$

$\Rightarrow A^T A = I$

$Ax = \lambda x$

$\Rightarrow (Ax)^T (Ax) = (\lambda x)^T (\lambda x)$

$\Rightarrow x^T A^T A x = |\lambda|^2 x^T x$

$\Rightarrow x^T x = |\lambda|^2 x^T x$ $\because$ eigenvectors are normalized

$\Rightarrow x^T x = 1$

$\Rightarrow |\lambda|^2 = 1$ #

(iii) Assume λ_1, λ_2 are the eigenvalues of A

x_1, x_2 are the eigenvectors of A

$\lambda_1 \neq \lambda_2$

$Ax_1 = \lambda_1 x_1, \quad Ax_2 = \lambda_2 x_2$

$(Ax_1)^H (Ax_2) = (\lambda_1 x_1)^H (\lambda_2 x_2)$

$\Rightarrow x_1^H A^H A x_2 = \bar{\lambda}_1 \lambda_2 x_1^H x_2$

($\because A^H A = I$)

$\Rightarrow x_1^H x_2 = \bar{\lambda}_1 \lambda_2 x_1^H x_2$

$\Rightarrow (\bar{\lambda}_1 \lambda_2 - 1) x_1^H x_2 = 0$

From (ii), $\bar{\lambda}_1 \lambda_2 = 1$ only if $\lambda_1 = \lambda_2$

$\Rightarrow x_1^H x_2 = 0$

it means the eigenvectors are orthogonal.

#

(iv)

Vector x will be rotated or reflected,

but the length will not be changed. #

(b)

(i) The left singular vectors of A are the eigenvectors of AA^T.

The right singular vectors of A are the eigenvectors of A^TA.

Singular Value Decomposition:

$$A = U\Sigma V^T$$

$$\begin{aligned} AA^T &= (U\Sigma V^T)(U\Sigma V^T)^T \\ &= U\Sigma V^T V \Sigma^T U^T \quad \left(\because \begin{array}{l} V^TV = I \\ \Sigma^T = \Sigma \end{array}\right) \\ &= U\Sigma^2 U^T \end{aligned}$$

$$\begin{aligned} A^TA &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U \Sigma V^T \quad \left(\because \begin{array}{l} U^TU = I \\ \Sigma^T = \Sigma \end{array}\right) \\ &= V\Sigma^2 V^T \end{aligned}$$

#

(ii) The singular values of A are the square root of the eigenvalues of AA^T and A^TA.

(c)

(i) False. ex: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ eigenvalues are all 1.

(ii) False. ex: the eigenvectors of $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$,

but their sum $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is not an eigenvector.

(iii) True. $Ax = \lambda x$

$\Rightarrow x^T A x = x^T(\lambda x) = x^T x \lambda \geq 0$

$\because x^T x$ is definitely non-negative

$\Rightarrow \lambda$ is also non-negative.

(iv) True. ex: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ → rank: 2

→ has only one distinct eigenvalue 1

(v) True. ex: the eigenvectors of $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ corresponding to the same eigenvalue 1.

Their sum $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is also an eigenvector.

2. (a) H denotes Head, T denotes Tail

(i)

$$P(H50 \mid T) = \frac{P(T \mid H50) \cdot P(H50)}{P(H50, T) + P(H60, T)}$$

$$= \frac{P(T \mid H50) \cdot P(H50)}{P(T \mid H50) \cdot P(H50) + P(T \mid H60) \cdot P(H60)}$$

$$= \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 0.4 \cdot 0.5}$$

$$= \frac{25}{45}$$

$$= \frac{5}{9} \approx 0.5556 \quad \#$$

(ii)

$$P(H50 \mid THHH) = \frac{P(THHH \mid H50) \cdot P(H50)}{P(H50, THHH) + P(H60, THHH)}$$

$$= \frac{P(THHH \mid H50) \cdot P(H50)}{P(THHH \mid H50) \cdot P(H50) + P(THHH \mid H60) \cdot P(H60)}$$

$$= \frac{(0.5)^4 \cdot 0.5}{(0.5)^4 \cdot 0.5 + (0.4 \cdot (0.6)^3) \cdot 0.5}$$

$$= \frac{3125}{3125 + 4320}$$

$$= \frac{625}{1489} \approx 0.4197 \quad \#$$

(iii) Let A denotes 9 Heads and 1 Tail in 10 flips, regardless the order.

①

$$P(H50|A) = \frac{P(A|H50)\cdot P(H50)}{P(H50,A)+P(H55,A)+P(H60,A)}$$

$$= \frac{P(A|H50)\cdot P(H50)}{P(A|H50)\cdot P(H50)+P(A|H55)\cdot P(H55)+P(A|H60)\cdot P(H60)}$$

$$= \frac{((0.5)^9\cdot 0.5)\cdot \frac{1}{3}}{((0.5)^9\cdot 0.5)\cdot \frac{1}{3}+((0.55)^9\cdot 0.45)\cdot \frac{1}{3}+((0.6)^9\cdot 0.4)\cdot \frac{1}{3}}$$

≈ 0.1379 #

②

$$P(H55|A) = \frac{P(A|H55)\cdot P(H55)}{P(H50,A)+P(H55,A)+P(H60,A)}$$

$$= \frac{P(A|H55)\cdot P(H55)}{P(A|H50)\cdot P(H50)+P(A|H55)\cdot P(H55)+P(A|H60)\cdot P(H60)}$$

$$= \frac{((0.55)^9\cdot 0.45)\cdot \frac{1}{3}}{((0.5)^9\cdot 0.5)\cdot \frac{1}{3}+((0.55)^9\cdot 0.45)\cdot \frac{1}{3}+((0.6)^9\cdot 0.4)\cdot \frac{1}{3}}$$

≈ 0.2927 #

③

$$P(H60 \mid A) = \frac{P(A \mid H60) \cdot P(H60)}{P(H50, A) + P(H55, A) + P(H60, A)}$$

$$= \frac{P(A \mid H60) \cdot P(H60)}{P(A \mid H50) \cdot P(H50) + P(A \mid H55) \cdot P(H55) + P(A \mid H60) \cdot P(H60)}$$

$$= \frac{((0.6)^9 \cdot 0.4) \cdot \frac{1}{3}}{((0.5)^9 \cdot 0.5) \cdot \frac{1}{3} + ((0.55)^9 \cdot 0.45) \cdot \frac{1}{3} + ((0.6)^9 \cdot 0.4) \cdot \frac{1}{3}}$$

≈ 0.5694 #

(b)

$$P(Preg \mid Pos) = \frac{P(Pos \mid Preg) \cdot P(Preg)}{P(Preg, Pos) + P(NotPreg, Pos)}$$

$$= \frac{P(Pos \mid Preg) \cdot P(Preg)}{P(Pos \mid Preg) \cdot P(Preg) + P(Pos \mid NotPreg) \cdot P(NotPreg)}$$

$$= \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.1 \cdot 0.99}$$

$= \frac{1}{11} \approx 0.0909$ #

If a woman received a positive test, only 9% probability that she is pregnant. It is a bad test. Because a large proportion (99%) of the female population is not pregnant, but the test returns "positive" 10%, it will cause many incorrect tests. #

(c)

$$E(x_i + b) = E(x_i) + b$$

$$\begin{aligned} E(Ax_i) &= E\left(\sum_{j=1}^{n} A_{i,j}\, x_j\right) \\ &= \sum_{j=1}^{n} A_{i,j}\, E(x_j) \\ &= \sum_{j=1}^{n} A_{i,j}\, E(x)_j \\ &= [A \cdot E(x)]_i \end{aligned}$$

$$\Rightarrow E(Ax) = A \cdot E(x)$$

$$\begin{aligned} \therefore\ E(Ax+b) &= E(Ax) + b \\ &= AE(x) + b \quad \# \end{aligned}$$

(d)

$$\mathrm{cov}(x) = E\left((x - E(x))(x - E(x))^T\right)$$

$$\mathrm{cov}(Ax+b) = E\left((Ax+b-E(Ax+b))(Ax+b-E(Ax+b))^T\right)$$

From (c) $= E\left((Ax+b-AE(x)-b)(Ax+b-AE(x)-b)^T\right)$

$$\begin{aligned} &= E\left((Ax - AE(x))(Ax - AE(x))^T\right) \\ &= E\left(A(x-E(x))(x-E(x))^T A^T\right) \\ &= A\, E\left((x-E(x))(x-E(x))^T\right) A^T \\ &= A\, \mathrm{cov}(x)\, A^T \quad \# \end{aligned}$$

3. (a)

$$\nabla_x x^T A y = A y \quad \#$$

(b)

$$\nabla_y x^T A y = A^T x \quad \#$$

(c)

$$\nabla_A x^T A y = \begin{bmatrix} \frac{\partial x^T A y}{\partial a_{1,1}} & \cdots & \frac{\partial x^T A y}{\partial a_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial x^T A y}{\partial a_{n,1}} & \cdots & \frac{\partial x^T A y}{\partial a_{n,m}} \end{bmatrix}$$

$$= x y^T \quad \#$$

(d)

$$\nabla_x f = \nabla_x (x^T A x + b^T x)$$

$$= \nabla_x (x^T A x) + \nabla_x (b^T x)$$

$$= A x + A^T x + b \quad \#$$

(e)

Assume $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times m}$

$$tr(AB) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j} b_{j,i}$$

$$\frac{\partial tr(AB)}{\partial a_{i,j}} = b_{j,i}$$

$$\nabla_A f = \nabla_A tr(AB) = B^T \quad \#$$

4.

$$\frac{1}{2}\sum_{i=1}^{n} \| y^{(i)} - Wx^{(i)} \|^2$$

$$= \frac{1}{2}\sum_{i=1}^{n} (y^{(i)} - Wx^{(i)})^T (y^{(i)} - Wx^{(i)})$$

$$= \frac{1}{2}\sum_{i=1}^{n} \left(y^{(i)T} y^{(i)} - y^{(i)T} Wx^{(i)} - (Wx^{(i)})^T y^{(i)} + (Wx^{(i)})^T (Wx^{(i)}) \right)$$

$$= \frac{1}{2}\sum_{i=1}^{n} \left(y^{(i)T} y^{(i)} - y^{(i)T} Wx^{(i)} - y^{(i)T} (Wx^{(i)}) + (Wx^{(i)})^T (Wx^{(i)}) \right)$$

$$= \frac{1}{2}\sum_{i=1}^{n} \left(y^{(i)T} y(i) - 2 y^{(i)T} \underset{\sim}{W} x^{(i)} + x^{(i)T} W^T \underset{\sim}{W} x^{(i)} \right)$$

Only take out the terms with W

$$\Rightarrow f(W) = \frac{1}{2}\sum_{i=1}^{n} \left(-2 y^{(i)T} W x^{(i)} + x^{(i)T} W^T W x^{(i)} \right)$$

$$= \sum_{i=1}^{n} \left(-y^{(i)T} W x^{(i)} + \frac{1}{2} x^{(i)T} W^T W x^{(i)} \right)$$

$$= \sum_{i=1}^{n} \left(-tr(y^{(i)T} W x^{(i)}) + \frac{1}{2} tr(x^{(i)T} W^T W x^{(i)}) \right)$$

$$= \sum_{i=1}^{n} \left(-tr(W x^{(i)} y^{(i)T}) + \frac{1}{2} tr(W x^{(i)} x^{(i)T} W^T) \right)$$

$$= -tr\left(W \sum_{i=1}^{n} x^{(i)} y^{(i)T}\right) + \frac{1}{2} tr\left(W \sum_{i=1}^{n} (x^{(i)} x^{(i)T}) W^T\right)$$

$$= -tr(WXY^T) + \frac{1}{2} tr(WXX^TW^T)$$

From Hint, $\frac{\partial f(W)}{\partial W}$

$$= -(XY^T)^T + \frac{1}{2}\left(W(XX^T)^T + W(XX^T)\right)$$

$$= -YX^T + \frac{1}{2}(WXX^T + WXX^T)$$

$$= -YX^T + WXX^T = 0$$

$$\Rightarrow WXX^T = YX^T$$

$$\Rightarrow W = YX^T(XX^T)^{-1}$$

$$= YX^{-1} \quad \#$$

Linear regression workbook

This workbook will walk you through a linear regression example. It will provide familiarity with Jupyter Notebook and Python. Please print (to pdf) a completed version of this workbook for submission with HW #1.

ECE C147/C247 Winter Quarter 2022, Prof. J.C. Kao, TAs Y. Li, P. Lu, T. Monsoor, T. wang

In [1]:

```
import numpy as np
import matplotlib.pyplot as plt

#allows matlab plots to be generated in line
%matplotlib inline
```

Data generation

For any example, we first have to generate some appropriate data to use. The following cell generates data according to the model: $y = x - 2x^2 + x^3 + \epsilon$

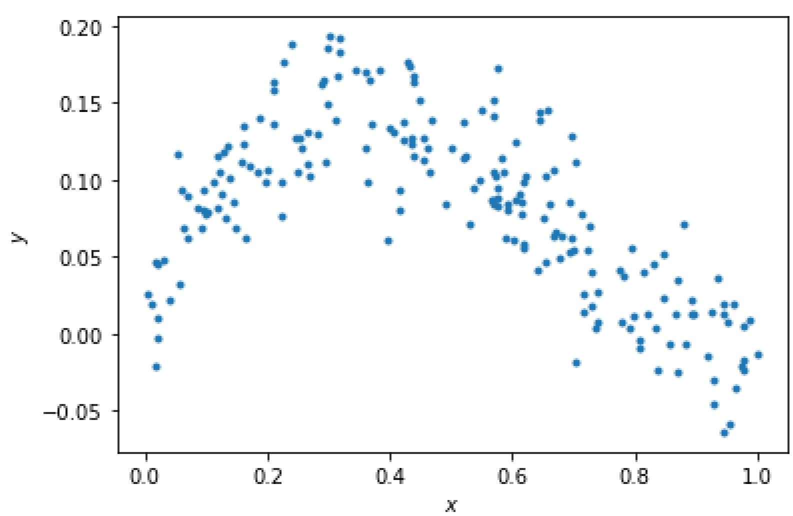
In [2]:

```
np.random.seed(0)  # Sets the random seed.
num_train = 200     # Number of training data points

# Generate the training data
x = np.random.uniform(low=0, high=1, size=(num_train,))
y = x - 2*x**2 + x**3 + np.random.normal(loc=0, scale=0.03, size=(num_train,))
f = plt.figure()
ax = f.gca()
ax.plot(x, y, '.')
ax.set_xlabel('$x$')
ax.set_ylabel('$y$')
```

Out[2]:

```
Text(0, 0.5, '$y$')
```



QUESTIONS:

Write your answers in the markdown cell below this one:

(1) What is the generating distribution of x?

(2) What is the distribution of the additive noise ϵ?

ANSWERS:

(1) It is the Uniform Distribution from 0 to 1.

(2) It is the Normal Distribution (mean=0 and standard deviation=0.03).

Fitting data to the model (5 points)

Here, we'll do linear regression to fit the parameters of a model $y = ax + b$.

In [3]:

```
# xhat = (x, 1)
xhat = np.vstack((x, np.ones_like(x)))

# ==================== #
# START YOUR CODE HERE #
# ==================== #
# GOAL: create a variable theta; theta is a numpy array whose elements are [a, b]

theta = np.matmul(np.matmul(np.linalg.inv(np.matmul(xhat, xhat.T)), xhat), y)
print(theta)

# ================== #
# END YOUR CODE HERE #
# ================== #
```

```
[-0.10599633  0.13315817]
```

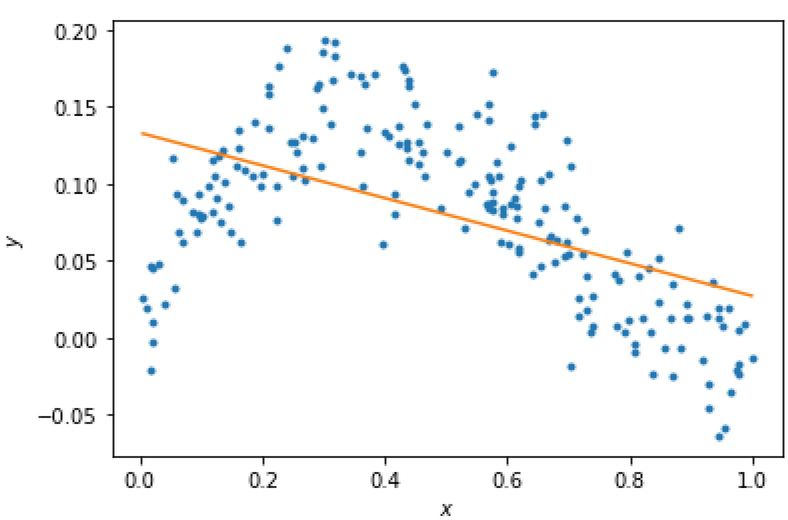
In [4]:

```
# Plot the data and your model fit.
f = plt.figure()
ax = f.gca()
ax.plot(x, y, '.')
ax.set_xlabel('$x$')
ax.set_ylabel('$y$')

# Plot the regression line
xs = np.linspace(min(x), max(x),50)
xs = np.vstack((xs, np.ones_like(xs)))
plt.plot(xs[0,:], theta.dot(xs))
```

Out[4]:

```
[<matplotlib.lines.Line2D at 0x7fd28a2547f0>]
```



QUESTIONS

(1) Does the linear model under- or overfit the data?

(2) How to change the model to improve the fitting?

ANSWERS

(1) It underfits the data.

(2) Increase the order of polynomial models.

Fitting data to the model (10 points)

Here, we'll now do regression to polynomial models of orders 1 to 5. Note, the order 1 model is the linear model you prior fit.

In [5]:

```
N = 5
xhats = []
thetas = []

# ==================== #
# START YOUR CODE HERE #
# ==================== #

# GOAL: create a variable thetas.
# thetas is a list, where theta[i] are the model parameters for the polynomial fit of orde
#   i.e., thetas[0] is equivalent to theta above.
#   i.e., thetas[1] should be a length 3 np.array with the coefficients of the x^2, x, and
#   ... etc.

for i in np.arange(N):
    if i == 0:
        xhat = np.vstack((x, np.ones_like(x)))
    else:
        xhat = np.vstack(((x**(i+1)), xhat))
    xhats.append(xhat)

for xhat in xhats:
    thetas.append(np.matmul(np.matmul(np.linalg.inv(np.matmul(xhat, xhat.T)), xhat), y))
print(thetas)

# ================== #
# END YOUR CODE HERE #
# ================== #
```

```
[array([-0.10599633,  0.13315817]), array([-0.48023061,  0.36743967,  0.05521084]), array
([ 0.8843808 , -1.82077417,  0.91178032,  0.00979068]), array([ 0.14080037,  0.60466289, -
1.64250929,  0.87250485,  0.01175321]), array([ 0.52432591, -1.164568  ,  1.76052438, -2.0
7430275,  0.93373916,
        0.009716  ])]
```

In [6]:

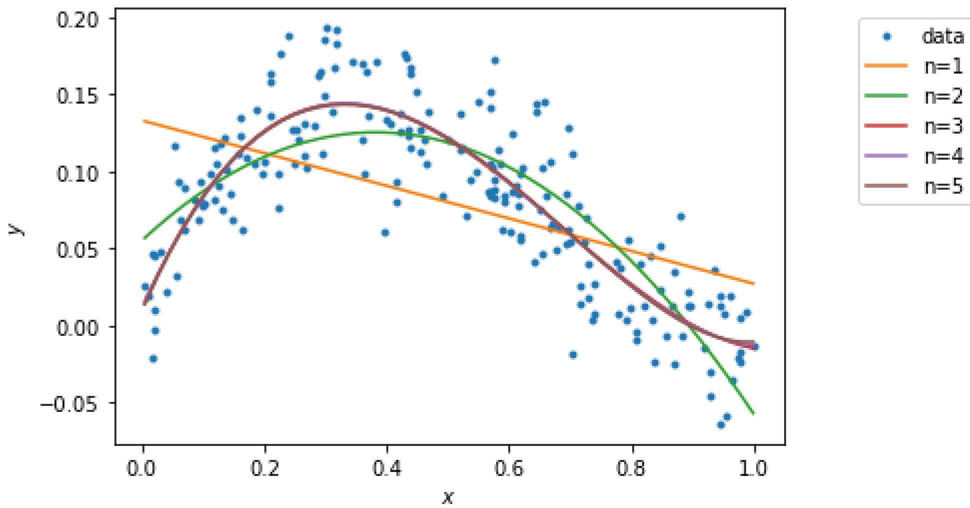
```
# Plot the data
f = plt.figure()
ax = f.gca()
ax.plot(x, y, '.')
ax.set_xlabel('$x$')
ax.set_ylabel('$y$')

# Plot the regression lines
plot_xs = []
for i in np.arange(N):
    if i == 0:
```

```
        plot_x = np.vstack((np.linspace(min(x), max(x),50), np.ones(50)))
    else:
        plot_x = np.vstack((plot_x[-2]**(i+1), plot_x))
    plot_xs.append(plot_x)

for i in np.arange(N):
    ax.plot(plot_xs[i][-2,:], thetas[i].dot(plot_xs[i]))

labels = ['data']
[labels.append('n={}'.format(i+1)) for i in np.arange(N)]
bbox_to_anchor=(1.3, 1)
lgd = ax.legend(labels, bbox_to_anchor=bbox_to_anchor)
```



Calculating the training error (10 points)

Here, we'll now calculate the training error of polynomial models of orders 1 to 5:

$L(\theta) = \frac{1}{2} \sum_j (\hat{y}_j - y_j)^2$

In [7]:

```
training_errors = []

# ==================== #
# START YOUR CODE HERE #
# ==================== #

# GOAL: create a variable training_errors, a list of 5 elements,
# where training_errors[i] are the training loss for the polynomial fit of order i+1.

for i in np.arange(N):
    training_errors.append(1/2 * np.sum((np.matmul(thetas[i], xhats[i]) - y)**2))

# ================== #
# END YOUR CODE HERE #
# ================== #

print ('Training errors are: \n', training_errors)
```

```
Training errors are: 
 [0.2379961088362701, 0.10924922209268528, 0.08169603801105374, 0.08165353735296979, 0.081
61479195525295]
```

QUESTIONS

(1) Which polynomial model has the best training error?

(2) Why is this expected?

ANSWERS

(1) Polynomial model with order 5.

(2) The higher degree polynomial model will have better training error because it tries to pass through more data points.

Generating new samples and validation error (5 points)

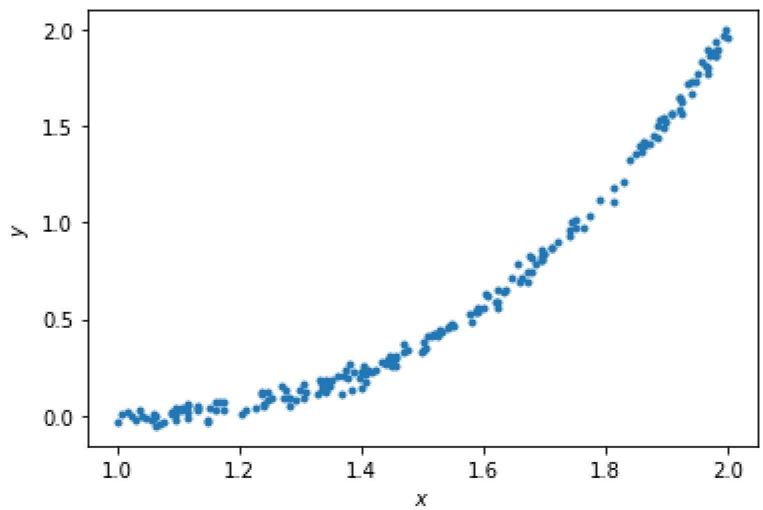
Here, we'll now generate new samples and calculate the validation error of polynomial models of orders 1 to 5.

In [8]:

```
x = np.random.uniform(low=1, high=2, size=(num_train,))
y = x - 2*x**2 + x**3 + np.random.normal(loc=0, scale=0.03, size=(num_train,))
f = plt.figure()
ax = f.gca()
ax.plot(x, y, '.')
ax.set_xlabel('$x$')
ax.set_ylabel('$y$')
```

Out[8]:

```
Text(0, 0.5, '$y$')
```



In [9]:

```
xhats = []
for i in np.arange(N):
    if i == 0:
        xhat = np.vstack((x, np.ones_like(x)))
        plot_x = np.vstack((np.linspace(min(x), max(x),50), np.ones(50)))
    else:
        xhat = np.vstack((x**(i+1), xhat))
        plot_x = np.vstack((plot_x[-2]**(i+1), plot_x))

    xhats.append(xhat)
```

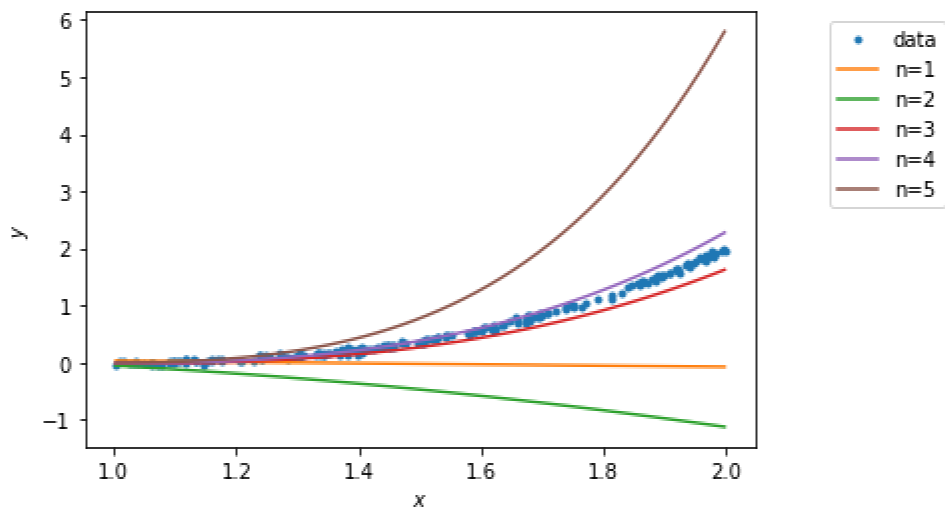
In [10]:

```
# Plot the data
f = plt.figure()
ax = f.gca()
ax.plot(x, y, '.')
ax.set_xlabel('$x$')
ax.set_ylabel('$y$')
```

```
# Plot the regression lines
plot_xs = []
for i in np.arange(N):
    if i == 0:
        plot_x = np.vstack((np.linspace(min(x), max(x),50), np.ones(50)))
    else:
        plot_x = np.vstack((plot_x[-2]**(i+1), plot_x))
    plot_xs.append(plot_x)

for i in np.arange(N):
    ax.plot(plot_xs[i][-2,:], thetas[i].dot(plot_xs[i]))

labels = ['data']
[labels.append('n={}'.format(i+1)) for i in np.arange(N)]
bbox_to_anchor=(1.3, 1)
lgd = ax.legend(labels, bbox_to_anchor=bbox_to_anchor)
```



In [11]:

```
validation_errors = []

# ==================== #
# START YOUR CODE HERE #
# ==================== #

# GOAL: create a variable validation_errors, a list of 5 elements,
# where validation_errors[i] are the validation loss for the polynomial fit of order i+1.

for i in np.arange(N):
    validation_errors.append(1/2 * np.sum((np.matmul(thetas[i], xhats[i]) - y)**2))

# ================== #
# END YOUR CODE HERE #
# ================== #

print ('Validation errors are: \n', validation_errors)
```

```
Validation errors are:
 [80.86165184550586, 213.19192445058104, 3.1256971084103693, 1.1870765210044922, 214.91021
752914227]
```

QUESTIONS

(1) Which polynomial model has the best validation error?

(2) Why does the order-5 polynomial model not generalize well?

ANSWERS

(1) Polynomial model with order 4.

(2) It is a overfitting problem because we generate a new set of data. The more complex model may not generalize well if the data come from a different dataset.