Mahcine Learning, Spring 2021
Homework 1

Daniil Bulat, Jonas Josef Huwyler, Haochen Li,Giovanni Magagnin

March 3, 2021

Exercise 1

Autonomous weapons are not operated with human involvement but utilize artificial intelligence instead. Autonomous weapons may present dangers:

The first concern is that, because of the plausible AI arms race, autonomous weapons could become as common as Kalashnikovs are today. With an army of autonomous weapons, the threshold to start a war might be lower, as theoretically fewer lives would be at risk as AI weapons compete against each other. Wars could therefore become more frequent in the future.

The second concern is that, unlike nuclear weapons, they are relatively cheap and could last a long time with few maintenance costs. Autonomous weapons could easily fall into the wrong hands and become harmful to society. Autonomous weapons, for example, could be ideal for assassinations, destabilizing nations, subjugating populations, and selectively killing a particular ethnic group. This could lead to an increase in terror activities in the future.

Other dangers that machine learning / artificial intelligence can present are the followings:

Threats to individuals: Sophisticated machine-driven algorithms could be developed to exploit phishing scams. The target would not be able to distinguish whether they are dealing with a human or a fraudulent entity. Financial information, digital identity, and private data could be easily stolen.

Threats to the economy: Automated algorithms trading in financial markets could increase volatility, instability, and unintended systematic risks. They could also crowd out less experienced investors.

Dangers to political stability: Democratic elections could be influenced by false ad hoc campaigns and messages. AI could be used for spreading political fundamentalism through targeted manipulation.

Dangers of job loss: Due to the automation fostered by the development of AI, workers performing repetitive and predictable tasks could soon be replaced by AI. Since such work is often done by lower social classes, this could lead to massive job loss and social injustice. How will states and politics deal with this transition to prevent an uproar of social injustice as well as rejection towards AI.

Exerceise 2

Have a look at the file *lin_reg_iris_first_steps.R1* on StudyNet (Course Resources) and make sure you understand all commands used in there. For comparability of the results please add the line *set.seed(123)* in the beginning of the file. This will make the sample command return the same results every time you use it.

The following is the basic setup for solving the exercise.

```r
# Set up
rm(list = ls())
set.seed(123)

# Shuffle the Iris data
Data = iris[sample(1:150),]

# and split into training and test data (80-20)
Data.Train = Data[1:120,]
Data.Test = Data[121:150,]
```

Qestion 1:

Determine the linear regression function in the form f(x1; x2) = m1x1 + m2x2 + c for predicting Sepal.Length depending on x1 =Petal.Length and x2 =Petal.Width on the training data.

The result below shows that the linear regression function for predicting would be

$$Sepal.Length = 4.17 + 0.54 * Petal.Length - 0.31 * Petal.Width$$

```r
# Regression on the training data - Model 1: f(x1; x2) = m1x1 + m2x2 + c
Iris.Model1 = lm(Sepal.Length ~ Petal.Length + Petal.Width, data = Data.Train)
summary(Iris.Model1)
```

```
##
## Call:
## lm(formula = Sepal.Length ~ Petal.Length + Petal.Width, data = Data.Train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17814 -0.32412 -0.04138  0.28352  1.04296
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.16844    0.10726  38.863  < 2e-16 ***
## Petal.Length  0.54270    0.07468   7.267 4.46e-11 ***
## Petal.Width  -0.31322    0.17293  -1.811   0.0727 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4108 on 117 degrees of freedom
## Multiple R-squared:  0.7692, Adjusted R-squared:  0.7652
## F-statistic: 194.9 on 2 and 117 DF,  p-value: < 2.2e-16
```

Question 2:

Do the same for only one attribute, x1 =Petal.Length on the training data.

The result below shows that the linear regression function for predicting would be

$$Sepal.Length = 4.28 + 0.41 * Petal.Length$$

```
# Regression on the training data - Model 2: f(x1) = m1x1 + c
Iris.Model2 = lm(Sepal.Length ~ Petal.Length, data = Data.Train)
summary(Iris.Model2)
```

```
##
## Call:
## lm(formula = Sepal.Length ~ Petal.Length, data = Data.Train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23656 -0.30673 -0.03334  0.26958  1.02598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.27856    0.08921   47.96   <2e-16 ***
## Petal.Length  0.41289    0.02120   19.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4148 on 118 degrees of freedom
## Multiple R-squared:  0.7627, Adjusted R-squared:  0.7607
## F-statistic: 379.2 on 1 and 118 DF,  p-value: < 2.2e-16
```

Question 3:

Do the same for the three attributes, x1 =Petal.Length, x2 =Petal.Width, x3 =Sepal.Width on the training data.

The result below shows that the linear regression function for predicting would be

$$Sepal.Length = 1.90 + 0.70 * Petal.Length - 0.53 * Petal.Width + 0.64 * Sepal.Width$$

```
# Regression on the training data - Model 3: f(x1; x2; x3) = m1x1 + m2x2 + m3x3 + c
Iris.Model3 = lm(Sepal.Length ~ Petal.Length + Petal.Width + Sepal.Width, data = Data.Train)
summary(Iris.Model3)
```

```
##
## Call:
## lm(formula = Sepal.Length ~ Petal.Length + Petal.Width + Sepal.Width,
##     data = Data.Train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82577 -0.21712  0.02843  0.18999  0.85864
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.89882    0.28429   6.679 8.74e-10 ***
## Petal.Length  0.69826    0.06208  11.247  < 2e-16 ***
## Petal.Width  -0.53303    0.13964  -3.817 0.000218 ***
## Sepal.Width   0.63637    0.07606   8.367 1.52e-13 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3258 on 116 degrees of freedom
## Multiple R-squared:  0.856,  Adjusted R-squared:  0.8523
## F-statistic: 229.9 on 3 and 116 DF,  p-value: < 2.2e-16
```

Question 4:

Find the commands for mean and variance in R and compute the mean and the variance of Petal.Length and Petal.Width, respectively.

By using the commands from R, the mean of variance can be shown as following:

	Mean	Variance
Petal.Length	3.81	3.22
Petal.Width	1.22	0.60

```
mean(Data.Train$Petal.Length)
```

```
## [1] 3.81
```

```
var(Data.Train$Petal.Length)
```

```
## [1] 3.215866
```

```
mean(Data.Train$Petal.Width)
```

```
## [1] 1.2275
```

```
var(Data.Train$Petal.Width)
```

```
## [1] 0.5998256
```

Bonus Question:

Use the mean and variance commands to compute (or verify) the regression function for part 2) step by step without the lm-command, using the formula for simple linear regression.

Using the formula to solve the minimization problem of simple linear regression:

$$\hat{\alpha} = \bar{y} - (\hat{\beta}\bar{x})$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{Cov(x, y)}{Var(x)}$$

The regression coefficients calculated by mean and variance are identical to the lm-command:

$$Sepal.Length = 4.28 + 0.41 * Petal.Length$$

```
beta.hat = var(Data.Train$Sepal.Length,Data.Train$Petal.Length)/var(Data.Train$Petal.Length)
alpha.hat = mean(Data.Train$Sepal.Length) - beta.hat*mean(Data.Train$Petal.Length)
```

```
alpha.hat
```

```
## [1] 4.278556
```

```r
beta.hat
```

```
## [1] 0.4128899
```

```r
Iris.Model2$coefficients
```

```
##  (Intercept) Petal.Length
##    4.2785562    0.4128899
```