Statistical Modeling and Machine Learning
01 - Math refresher

Julien Gagneur, Žiga Avsec, Leonhard Wachutka, Johannes Söding

April 25, 2017

1 Reading

1.1 Maths

You can find a concise summary of the required math background in appendix A.1-A.2 of the freely available book by Barber: http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.Online. (Appendix A.3 to A.6 can also become helpful later.). Alternatively, you can use Chapter 2 up to section 2.11 determinant included from http://deeplearningbook.org/contents/linea_algebra.html

Online lectures, exercises and exam questions for linear algebra can be found here: http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/

Relevant videos are also available at the coursera machine learning lecture (Section III. Linear Algebra Review): https://class.coursera.org/ml-005/lecture

1.2 R programming

Get familiar with R basics by going through the (very) short introduction to R by Paul Torfs and Claudia Brauer. http://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf. The TODO sections in this tutorial are recommended but you do not need to hand in the solutions.

2 Calculating with matrices and gradients

Question 1: Let $\mathbf{A} \in R^{m \times n}$ and $\mathbf{B} \in R^{n \times l}$ be matrices. Show by writing componentwise that

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T \tag{1}$$

Question 2: Show for invertible matrices $\mathbf{A}$ and $\mathbf{B}$ that

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{2}$$

by using the definition of $(\mathbf{AB})^{-1}$ as the matrix that gives the identity matrix when multiplied from left and right with $\mathbf{AB}$.

In the following, we will use an alternative notation for the gradient:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}. \tag{3}$$

Question 3: Let $\mathbf{x}$ and $\mathbf{b}$ be vectors, $\mathbf{A}$ a matrix. Show componentwise that

$$\frac{\partial(\mathbf{b}^T\mathbf{x})}{\partial \mathbf{x}} = \mathbf{b} \tag{4}$$

Question 4: Let $\mathbf{A}$ be a symmetric matrix. Show by componentwise partial derivatives that

$$\frac{\partial}{\partial \mathbf{x}}\left(\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = \mathbf{A}\mathbf{x} \tag{5}$$

Question 5: Let $\mathbf{x}$ be a vector in R^D, $\mathbf{A}$ a matrix in $R^{D\times D}$. Show the following equations by writing them componentwise:

Show that

$$\nabla_{\mathbf{x}}||\mathbf{x}|| = \mathbf{x}/||\mathbf{x}|| \tag{6}$$

Question 6: Show that

$$\mathbf{P} = \frac{aa^T}{||a||^2} \tag{7}$$

is the projection operator onto the one-dimensional vector space spanned by $\mathbf{a}$. Hint: Any vector $\mathbf{x} \in R^D$ can be split into two parts: the part within $\mathrm{span}(\mathbf{a})$, $\alpha\mathbf{a}$, and the part orthogonal to it, $\mathbf{x}_\perp$.

3 Optimization

Question 7: In R, plot the function $f(x) = x\exp(-\frac{(x-1)^2}{2})$ for 100 points in $[-2, 2]$ using `x=seq(-2,2, length.out=100)`. How many local extrema do you see? Compute analytically their values and marke the solution by vertical lines using the function `abline(v=...)`.

4 Data exploration in R

The famous (Fisher's or Anderson's) iris data set (https://en.wikipedia.org/wiki/Iris_flower_data_set) gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica. The data is stored as a data.frame, which is a matrix like object. The main advantage is the possibility to store multiple different types together, whereas a matrix can only store one type (e.g integer). This is possible, because the object is internally stored as a list of elements of equal length. Some essential functions to explore and view the data are:

`str()`: Gives the structure of an object

`head()/tail()`: to view the first/last six rows of the dataset

`myData$myColumn`: Access the column of a data.frame

`myData[rowvector, columnvector]`: access specific rows and columns of the data.frame by number

`myData[rowname, columnname]`: access specific rows and columns of the data.frame by name

`myData$newColumn = value`: add a new column (must be of same length as the other)

`?myObject`: Get help for an object

```
data("iris")
str(iris)

## 'data.frame': 150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa

tail(iris)

##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 145          6.7         3.3          5.7         2.5 virginica
## 146          6.7         3.0          5.2         2.3 virginica
## 147          6.3         2.5          5.0         1.9 virginica
## 148          6.5         3.0          5.2         2.0 virginica
```

```
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
```

```
head(iris$Species)
```

```
## [1] setosa setosa setosa setosa setosa setosa
## Levels: setosa versicolor virginica
```

```
iris[1:10,2]
```

```
##  [1] 3.5 3.0 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1
```

```
head(iris[, "Sepal.Width"])
```

```
## [1] 3.5 3.0 3.2 3.1 3.6 3.9
```

Note that, the column "Species" is ob class Factor, which appears like a string/character object but differs significantly, as it internally stores all levels (in this case three) as integers and associates each level with a label.

```
chrObject = as.character(iris$Species)
str(iris$Species)
```

```
##  Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
str(chrObject)
```

```
##  chr [1:150] "setosa" "setosa" "setosa" "setosa" "setosa" ...
```

```
levels(iris$Species)
```

```
## [1] "setosa"     "versicolor" "virginica"
```

```
str(as.numeric(iris$Species))
```

```
##  num [1:150] 1 1 1 1 1 1 1 1 1 1 ...
```

```
str(as.numeric(chrObject))
```

```
## Warning in str(as.numeric(chrObject)):  NAs introduced by coercion
```

```
##  num [1:150] NA NA NA NA NA NA NA NA NA NA ...
```

Question 8: Given the "iris" dataset, answer the following questions: For each variable (except Species), what is the mean, the median, the interquartile range (range of the middle 50% the data) and the entire range of the data?

Hint: Five-number summary

Question 9: Plot a histogram of Sepal.Length with the default breaks and with 20 breaks.

Hint: ?hist

Question 10: How many of each of the flower species have a sepal length smaller than 5.5 and how many more?

Hint: New groups can be easily attached as new columns and ?table

Question 11: Plot Sepal.Length against Petal.Length, the finding from 2.), every variable against each other and all the numeric variables against each other.

Hint: ?plot and ?pairs

Question 12: Produce seperate plots of Petal.Width vs. Petal.Length for each Species combined in one plotting window.

Hint: ?par and mfrow argument

5 Projections in 3D

HOMEWORK (This is a homework excercise only and will not be covered in class. It is still due to submission.)

Define a pyramid as a scaffold of points along its edges:

```
p1= c(-1,-1,0)
p2= c(-1,1,0)
p3= c(1,1,0)
p4= c(1,-1,0)
p5= c(0,0,2)

v = seq(0,1,0.05)
e1 = v %o% (p2-p1) + rep(1,21) %o% p1
e2 = v %o% (p3-p2) + rep(1,21) %o% p2
e3 = v %o% (p4-p3) + rep(1,21) %o% p3
e4 = v %o% (p1-p4) + rep(1,21) %o% p4
e5 = v %o% (p5-p1) + rep(1,21) %o% p1
e6 = v %o% (p5-p2) + rep(1,21) %o% p2
e7 = v %o% (p5-p3) + rep(1,21) %o% p3
e8 = v %o% (p5-p4) + rep(1,21) %o% p4

pyramid = rbind(e1,e2,e3,e4,e5,e6,e7,e8)
```

Question 13: Write down the matrix that projects from (x, y, z) to (x, y) and plot the (x, y) points of the pyramid. Be careful about the order and dimensions of the matrices that you multiply.

Question 14: Now define two rotation matrices in two dimensions, R1 and R2, one rotating around the z axis by phi and the other rotating around the y axis by theta. Use them to rotate the pyramid and to project in onto the x-y plane. Try different angles phi and theta, e.g. `phi = 0.05*pi`, `theta = 0.3*pi`.

6 Reading

6.1 Assignment

Read Bishop's book up to section 1.2.4 (The Gaussian distribution) included. Optionnally, you can read sections 1.2.5-6.

6.2 Remarks

We will not look into reinforcement techniques.
We will look into the bootstrap later.
Figure 1.8 and Table 1.2 are not consistent, but the argumentation is right.
Equation (1.27): Note that g has to be a bijection to be a change of variable.
Equation (1.33): Note that f can be vector-valued.
Equation (1.37): We prefer the notation $\mathrm{E}_{X|Y}(f(X)) = \dots$

6.3 Questions

Question 15: **(At home only)** What is supervised learning and what is unsupervised learning? Give some examples.