Biostatistics [SBE304] (Fall 2019)
Project

Prof. Ayman M. Eldeib   Asem Alaa

Friday 18th October, 2019

1   Project phases

1. Announce your team (done)

2. Project proposal submission (25 October)

3. Feedback from TA on the proposal (26-28 October)

4. Submit project milestone (a simple working prototype) (25 November)

5. Feedback from TA on the milestone (26-30 November)

6. Submit final project and publicity (15 December)

7. Project presentation (TBA)

8. Announcing scores and the best projects (TBA)

View or add above phases to your calendar.

 Project submission repository is generated through `https://classroom.github.com/g/FIcbND8l`
You deliverables in this project are:

1. Project proposal: you will find a LaTeX template in your group repository.

2. Implementation in R: uploaded to the group repository.

3. Blog post documenting the project by each member at their personal website.

2   Project Proposal

To start writing your proposal, you must first have:

1. Data

2. Idea

2.1   Suggested list of projects

You can alternatively look for other interesting datasets for your proposal, but finally has to be reviewed and approved by the TA. The problem doesn't necessarily need to address health/biology related issues.

Data	Web	# Examples (n)	# Features (m)	Task (idea)
Hepatocellular Carcinoma (HCC) Sruvival	🔗	165	49	Binary classification: predict whether a patient dies within a year or survives.
Thyroid Disease	🔗	215	5	Multiclass classification: predict whether a patient given patient is normal (1) or suffers from hyperthyroidism (2) or hypothyroidism (3).
Haberman's Survival	🔗	306	3	Binary classification: predict whether a patient dies within five years or survives.
Thoracic Surgery Sruvival	🔗	470	17	Binary classification: predict whether a patient dies within a year after the surgery or survives
Echocardiogram Sruvival	🔗	132	12	Binary classification: predict whether a patient dies within a year or survives.
Chronic Kidney Disease	🔗	400	25	Binary classification: predict whether a patient has chronic kidney disease (CKD) or not.

2.2 Proposal required sections

1. Background and motivation; the reasons made you choose this project, for example it meets your research interests.

2. Project objectives

3. Data

4. Pre-processing; have you made any:

   - Feature selection
   - Feature normalization
   - Data imputation

   and why it was necessary.

5. Exploratory data analysis (EDA); some visualizations on the data before going further.

6. Methodology: you need to choose at least two methods for the classification task and compare the results.

7. Project schedule; you should provide a tentative plan for the whole project. The schedule might be in a form of weekly deadlines, for example a deadline for learning **Gaussian Naive Bayes** and applying it on standard dataset. The plan also may include the contribution of each team member in the project.

8. The personal websites of the team members. Expected to contain no content during the proposal submission.

2.3   The allowed methods

For the classification tasks, **you are only allowed to use methods from the following list**:

- Naive Bayes (NB) Classier (or Gaussian NB Classifier)

- Decision Trees

- K-nearest neighbors (KNN) model

- Logistic regression

- Linear Discriminant Analysis (LDA)

3   Publicity of your Project

In this project you are required to publish your project as blog on your personal website. The personal website should be updated at least with your resume/CV. In this blog you need to provide a storytelling for the project. It need to include the same components of the project proposal 2.2.

3.1   Personal website

Before launching your own personal website, you need first to learn some motivations on "why do you need to have a personal website" at the first place. The following list includes few short articles (5 minutes read each) that you are highly recommended to read before you start:

- "You should have a personal web site".

- "Why You Need a Personal Website".

- "7 Reasons Why You'll Regret Not Having A Personal Website".

Static website generators   Front-end side is all you care for a personal website. Such websites that comprises front-end without a back-end are called **static websites**. For static websites, you may build up your website using a single `index.html` file. But then you realize that you would like to add some look and pleasing behavior to your website, so you start adding additional front-end components like style-sheets (or `CSS` files) or additional libraries like `JQuery` and `bootstrap`.

   A very recommended and fast solution is using static website generators, that let you write all your content in a simpler mark-up language like `Markdown` then it handles all the front-end components on your behalf; very optimal solution for a personal website.

   Several static website generators exist on the internet including but not limited to: `Jekyll`, `Hugo`, and `Gatsby`. The short article "How to Choose a Static Site Generator" should guide you how to choose your static website generator.

To make the process of "starting your personal website" much more easier and faster, many developers upload templates for different purposes, **so all you need is to clone one template and start add your own information**. In this document, I list several templates dedicated for a personal website using `Jekyll` as a generator, since `Jekyll` is my favorite choice for several reasons like its wide support by GitHub pages that generates the website on the cloud, following each commit.

Example `Jekyll` templates for personal website

- `indigo` [source][live demo]

- `leonids` [source][live demo]

- `agusmakmun` [source][live demo]

- `modern-resume-theme` [source][live demo]

- `michaeldwan.com` [source][live demo]

- `jimeh.me` [source][live demo]

- `richgrundy.com` [source][live demo]

- `h313.info` [source][live demo]

- `simplest` [source][live demo]

- `Phantom` [source][live demo]

- `Online CV` [source][live demo]

- `Vitae` [source][live demo]

- `sustain` [source][live demo]

- `Chalk` [source][live demo]

- `resume-template` [source]

Hundreds of templates and themes can be found elsewhere

- `awesome-jekyll-websites`

- `jekyllthemes.io`

3.2 Hosting your website

GitHub offers you hosting your website using `github_username.github.io` for free. However, if you can prove your studentship, you can get more advantages from the GitHub Student Educational Pack, like a custom domain for free on `name.com` or `namecheap` .

4   Grading criteria & project checklist

☐ Project proposal that must include the minimum requirements.

☐ Project plan.

☐ Data pre-processing, when appropriate.

☐ Exploratory data analysis.

☐ Learning and classification using at least two different methods.

☐ Performance evaluation of classifiers and comparisons between at least two different methods.

☐ Implementation: code readability and authenticity.

☐ Project presentation.

☐ Project publicity: blog post by each team member on his personal website, and each personal website include a detailed CV/resume.