



## Reinaldo (Rei) Sanchez-Arias

Assistant Professor of Data  
Science

[Florida Polytechnic University](#)





## Biography

I am an Assistant Professor in the Department of [Data Science and Business Analytics](#) at Florida Polytechnic University ([Florida Poly](#)). Before joining Florida Poly, I served as the Program Director for the MS in [Big Data Analytics](#) in the School of Science at St Thomas University ([STU](#)) in Miami. I was also part of of the Applied Mathematics Department at Wentworth Institute of Technology ([WIT](#)) in Boston, and completed a postdoctoral researcher appointment for the Army High Performance Computing Research Center ([AHPCRC](#)) at The University of Texas at El Paso ([UTEP](#)) in collaboration with a group at Stanford University. Research projects for the AHPCRC involved global optimization and parameter estimation, reduced-order modeling, and data analytics.

## Interests

- Data Mining and Machine Learning
- Data Science Education
- Operations Research and Numerical Optimization

## Education

-  PhD in Computational Science, 2013  
The University of Texas at El Paso
-  MS in Computational Science, 2011  
The University of Texas at El Paso
-  BSc in Mathematics, 2008  
Universidad del Valle (Colombia)



## Why Jupyter Notebooks?



A Jupyter Notebook is an open-source web application to create and share documents that contain *live code*, *visualizations* and *narrative* text.

Jupyter notebooks are useful for data analysis, numerical simulation, statistical modeling, data visualization, machine learning, and much more.



File Edit View Insert Cell Kernel Widgets Help Snippets Not Trusted R O

Run Run C Markdown Keyboard Plot Appmode Validate

## Example 2

Let  $g(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$ , which is called **Himmelblau's function**.

To use the graph below:

- Hover the mouse over the interactive graph to view the point values
- Click and rotate the graph to view at different angles
- Hover over graph and scroll to zoom in and out

```
In [15]: g <- function(x1, y1) (x1^2 + y1 - 11)^2 + (x1 + y1^2 - 7)^2
x <- seq(-4.5, 4.5, by = .2)
y <- seq(-4.5, 4.5, by = .2)
z2 <- outer(x, y, g)
```

```
In [16]: # Creating an interactive plot that can be rotated to view points at different angles
myp <- plot_ly(z = ~z2) %>%
  add_surface()
myp
```



# Teaching with Jupyter Notebooks

Aim for and promote **reproducibility**

Use them in **class demos** and **homework** assignments

Notebooks can be **exported** as an .html file to share with others (e.g. students can easily upload them to your LMS)

Students get used to creating a *narrative*, explaining the *code* they use, and commenting on the *results*

## Models with `scikit-learn`

### Loading a data set

We will load in the `digits` data set that comes with `scikit-learn`

```
In [9]: # ----- Import `datasets` from `sklearn` -----  
from sklearn import datasets  
  
# Load in the `digits` data  
digits = datasets.load_digits()  
  
# Print the `digits` data  
#print(digits)
```

Note that the `datasets` module contains other methods to load and fetch popular reference datasets. You can check a list of available data sets and tools in the [scikit-learn website](#).

### Exploring the data set

You can access the `digits` data through the attribute `data`. Similarly, you can also access the target values or labels through the `target` attribute and the description through the `DESCR` attribute.

To see which keys you have available to already get to know your data, you can just run `digits.keys()`.



## Jupyter Notebooks in nanoHUB

<https://nanohub.org/tools/jupyter>

Free to create an account

Explore sample notebooks contributed by the community

Share notebooks with students and others

*Web-based tool:* just need a web-browser, no installation needed, everything runs remotely in nanoHUB resources

Collect

# Jupyter Notebook

Starts the Jupyter notebook server using the latest installed release of anaconda.

Launch Tool

Version 1.7 - published on 27 Jan 2020

doi:10.21981/W6TE-1750 [cite this](#)

🔒 This tool is closed source.

[View All Supporting Documents](#)

- 👤 [1313 users, detailed usage](#)
- 🗉 [0 Citation\(s\)](#)
- 💬 [2 questions \(Ask a question\)](#)
- ★ [0 review\(s\) \(Review this\)](#)
- 📌 [2 wish\(es\) \(New Wish\)](#)

→ Share: [f](#) [t](#) [s](#)...

- About
- Usage
- Citations
- Questions
- Reviews
- Wishlist
- Versions
- Supporting Docs

Category

[Tools](#)

Published on

27 Jan 2020

Watch resource

When watching a resource, you will be notified of changes made. You may stop watching at any time.

In [1]: `pip list`

Package	Version
absl-py	0.7.1
aflow	0.0.10
aiohttp	3.5.4
alabaster	0.7.12
alembic	1.0.11
amqp	2.4.2
anaconda-client	1.7.2
anaconda-navigator	1.9.6
anaconda-project	0.8.2
ann-visualizer	2.5
ansiwrap	0.8.4
antimony	2.11.0
appdirs	1.4.3
appmode	0.5.0
argh	0.26.2
arrow	0.13.1
ase	3.17.0
asn1crypto	0.24.0

In [ ]:



Jupyter LongExpOpt (autosaved) Terminate Session

View Insert Cell Kernel Widgets Help Snippets Trusted R O

Run Cell Markdown Code Snippets Validate

## # Optimization Problems in Statistical Learning: Some Examples

Below we describe some of the optimization problems that arise in machine learning problems. Notice that coming up with an objective function does not automatically solve the problem is just one of the steps in machine learning applications. Certain optimization problems are much harder than others, and choosing an appropriate method is critical. If you have big data, time complexity is a key factor to consider. Similarly, a solution should be not only possible but also feasible to obtain.

### ## Maximum Likelihood Estimation (MLE)

Let  $\mathbf{X}$  represent observed data and  $\theta$  represent the model parameters. If  $f(\mathbf{x}; \theta)$  is the probability density (mass) function, and  $L(\theta; \mathbf{X}) = \prod_{i=1}^n f(\mathbf{x}_i; \theta)$  is the likelihood function, then the maximum likelihood estimation seeks the solution of the problem:

$$\max_{\theta} L(\theta; \mathbf{X})$$

MLE is useful across many areas of [statistical inference](https://en.wikipedia.org/wiki/Statistical_inference) ([https://en.wikipedia.org/wiki/Statistical\\_inference](https://en.wikipedia.org/wiki/Statistical_inference)).

For example, MLE in [logistic regression](https://en.wikipedia.org/wiki/Logistic_regression) takes the form:

Given  $y_i \in \{-1, 1\}^n$ ,  $\mathbf{X} \in \mathbb{R}^{n \times p}$ , and  $\beta \in \mathbb{R}^p$ , the likelihood function is given by:

$$L(\beta) = \prod_{i=1}^n \Pr(y_i = 1; \mathbf{x}_i, \beta) = \prod_{i=1}^n \frac{1}{1 + e^{-y_i \mathbf{x}_i^T \beta}}$$

with the maximization of the *log-likelihood* function problem given by:

$$\max_{\beta} \ln L(\beta) = \max_{\beta} \sum_{i=1}^n \ln \left( \frac{1}{1 + e^{-y_i \mathbf{x}_i^T \beta}} \right)$$

(this is an unconstrained optimization problem)

nanoHUB Jupyter LongExpOpt (unsaved changes) Terminate Session

File Edit View Insert Cell Kernel Widgets Help Snippets Trusted R O

Code Snippets Snippets Validate

## Optimization Problems in Statistical Learning: Some Examples

Below we describe some of the optimization problems that arise in machine learning problems. Notice that coming up with an objective function does not automatically solve the problem is just one of the steps in machine learning applications. Certain optimization problems are much harder than others, and choosing an appropriate method is critical. If you have big data, time complexity is a key factor to consider. Similarly, a solution should be not only possible but also feasible to obtain.

### Maximum Likelihood Estimation (MLE)

Let  $\mathbf{X}$  represent observed data and  $\theta$  represent the model parameters. If  $f(\mathbf{x}; \theta)$  is the probability density (mass) function, and  $L(\theta; \mathbf{X}) = \prod_{i=1}^n f(\mathbf{x}_i; \theta)$  is the likelihood function, then the maximum likelihood estimation seeks the solution of the problem:

$$\max_{\theta} L(\theta; \mathbf{X})$$

MLE is useful across many areas of [statistical inference](https://en.wikipedia.org/wiki/Statistical_inference).

For example, MLE in [logistic regression](https://en.wikipedia.org/wiki/Logistic_regression) takes the form:

Given  $y \in \{-1, 1\}^n$ ,  $\mathbf{X} \in \mathbb{R}^{n \times p}$ , and  $\beta \in \mathbb{R}^p$ , the likelihood function is given by:

$$L(\beta) = \prod_{i=1}^n \Pr(y_i = 1; \mathbf{X}, \beta) = \prod_{i=1}^n \frac{1}{1 + e^{-y_i \mathbf{x}_i^T \beta}}$$

with the maximization of the *log-likelihood* function problem given by:

$$\max_{\beta} \ln L(\beta) = \max_{\beta} \sum_{i=1}^n \ln \left( \frac{1}{1 + e^{-y_i \mathbf{x}_i^T \beta}} \right)$$

(this is an unconstrained optimization problem)

Files **Running** Formgrader Courses Assignments

Select items to perform actions on them.

Upload New ↕

<input type="checkbox"/> 0	<a href="#">/ rjupyternb</a>	Name ↓	Last Modified	File size
<input type="checkbox"/>	..		seconds ago	
<input type="checkbox"/>	bin		6 days ago	
<input type="checkbox"/>	data		6 days ago	
<input type="checkbox"/>	doc		6 days ago	
<input type="checkbox"/>	examples		6 days ago	
<input type="checkbox"/>	figures		6 days ago	
<input type="checkbox"/>	middleware		6 days ago	
<input type="checkbox"/>	rappture		6 days ago	
<input type="checkbox"/>	rjupyternb		4 days ago	
<input type="checkbox"/>	src		6 days ago	
<input type="checkbox"/>	 LongExpOpt.ipynb		a minute ago	14.1 kB
<input type="checkbox"/>	 OptimApp1.ipynb		6 days ago	4.74 MB
<input type="checkbox"/>	 OptimApp2.ipynb		Running 6 days ago	13.7 kB
<input type="checkbox"/>	 Optimization.ipynb		6 days ago	9.58 MB
<input type="checkbox"/>	 RootApp1.ipynb		6 days ago	15.9 kB



## Some ideas for using nanoHUB in your course

Provide your students with *templates* (.ipynb) for class demos/HW

You could create a *group* in nanoHUB for your course: forum, announcements, wiki page, and easy file sharing

Share/Develop a *tool* with/for your students to explore topics of your course and test different implementations

Super helpful *Support Team*: contact them if your tool requires a special (or a particular version) library/package

# Rei Sanchez-Arias



+ Add Modules

## MY PROJECTS ⚙️ ✕

☰ All Projects + New Project

- 🔍 Data Mining Applications in...  
By me | manager
- 🔍 Scientific Computing and...  
By NCN URE/UCEE 2020 | collaborator
- 🔍 Tanya's personal project  
By Tanya Faltens | collaborator

## MY GROUPS ⚙️ ✕

Recent All

- 👤 nanoHUB Undergraduate Education Workshop  
approved
- 👤 Purdue Summer Undergraduate Research Symposium 2020  
approved

+ New Group

## MY TOOLS ✕

Recent Favorites All Tools

- Machine Learning for Materials Science: Part 1 📁
- Interactive Learning Tools for Scientific Computing and Data Analysis Using R 📁
- Jupyter Notebook 📁
- FIFA 2010 visualization 📁

## MY SESSIONS ⚙️ ✕

📄 Jupyter Notebook ⌵

- 🏠 Dashboard
- 👤 Profile
- ⚙️ Account
- 📁 Collections 1