

UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC)  
– BARCELONATECH

DEGREE FINAL PROJECT

---

# Using Random Fourier Features with Random Forest

---

*Author:*  
Albert RIBES

*Supervisor:*  
Lluís A. BELANCHE

Computer Science

March 28, 2019



UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC) – BARCELONATECH

# *Abstract*

Facultat d'Informàtica de Barcelona (FIB)  
Computer Science

Bachelor Degree in Computer Science

**Using Random Fourier Features with Random Forest**

by Albert RIBES

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

- En 3 idiomas



# Contents

<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem to solve . . . . .	1
1.2 Why is it important? . . . . .	1
1.3 Project proposal . . . . .	1
<b>2 Background Information and Theory</b>	<b>3</b>
2.1 Machine Learning . . . . .	3
2.2 Review de los principales modelos que existen . . . . .	3
2.2.1 Decision Tree . . . . .	3
2.2.2 Logistic Regression . . . . .	3
2.2.3 Support Vector Machines . . . . .	3
2.3 Ensemble Methods . . . . .	3
2.3.1 Bagging . . . . .	3
2.4 The kernel trick . . . . .	3
2.4.1 The RBF kernel . . . . .	3
2.5 Random Fourier Features . . . . .	3
2.6 Nystroem . . . . .	3
<b>3 Project Development</b>	<b>5</b>
3.1 General Idea . . . . .	5
3.1.1 State of the art con las RFF . . . . .	5
3.1.2 State of the art con las Nystroem . . . . .	5
3.2 Hyper-parameters . . . . .	5
3.3 Hypothesis . . . . .	5
3.3.1 Planteamiento de los experimentos . . . . .	5
3.4 Datasets . . . . .	5
<b>4 Experimental Results</b>	<b>7</b>
4.1 Enfrentar resultados 2 a 2 . . . . .	7
4.2 Contrastar hipótesis con resultados . . . . .	7
<b>5 Conclusion and Future Directions</b>	<b>9</b>
<b>6 Sustainability Report</b>	<b>11</b>
6.1 Environmental . . . . .	11
6.2 Economic . . . . .	11
6.3 Social . . . . .	11
6.3.1 Impacto Personal . . . . .	11
6.3.2 Impacto Social . . . . .	11
6.3.3 Riesgos Sociales . . . . .	11



# List of Figures





# List of Tables



# List of Abbreviations

<b>ML</b>	<b>Machine Learning</b>
<b>SVM</b>	<b>Support Vector Machine</b>
<b>RBF</b>	<b>Radial Basis Function</b>
<b>DT</b>	<b>Decision Tree</b>
<b>RF</b>	<b>Random Forest</b>
<b>RFF</b>	<b>Random Fourier Features</b>
<b>CV</b>	<b>Cross-Validation</b>



## Chapter 1

# Introduction

### 1.1 Problem to solve

- Trade-off between accuracy and train time is not good

### 1.2 Why is it important?

- Avances en este campo permitirían usarlo en otras ciencias como medicina, economía, sociedad
- Muchas tareas que ahora tiene que hacer un humano podría hacerlas una máquina, ahorrando tiempo y dinero

### 1.3 Project proposal

- Existe una batería de técnicas que son buenas, pero que nadie las ha combinado. Son:
  - Modelos simples
  - Ensembles
  - kernel trick
  - Aproximaciones de kernel
- La propuesta es combinar todo esto para mejorar el trade-off
- Sostenemos las siguientes hipótesis:
  - Se podría hacer un ensemble con modelos distintos a DT
  - Se puede aproximar una RBF-SVM pero con el coste de una lineal
  - RFF + Bootstrap quizá es demasiado aleatorio
  - Los modelos que no se basan en productos escalares no se beneficiarán tanto de usar RFF
- Lo que se hará en cada capítulo del trabajo



## Chapter 2

# Background Information and Theory

### 2.1 Machine Learning

- Clasificación y regresión
- Cross-validation
- Qué son los datos de train y test, y por qué se hace esa partición
- Qué es el sobre-ajuste

### 2.2 Review de los principales modelos que existen

#### 2.2.1 Decision Tree

#### 2.2.2 Logistic Regression

#### 2.2.3 Support Vector Machines

### 2.3 Ensemble Methods

#### 2.3.1 Bagging

- Bootstrap
- Random Forest

### 2.4 The kernel trick

#### 2.4.1 The RBF kernel

### 2.5 Random Fourier Features

### 2.6 Nystroem





## **Chapter 3**

# **Project Development**

### **3.1 General Idea**

#### **3.1.1 State of the art con las RFF**

#### **3.1.2 State of the art con las Nystroem**

### **3.2 Hyper-parameters**

### **3.3 Hypothesis**

#### **3.3.1 Planteamiento de los experimentos**

### **3.4 Datasets**



## **Chapter 4**

# **Experimental Results**

**4.1 Enfrentar resultados 2 a 2**

**4.2 Contrastar hipótesis con resultados**



## **Chapter 5**

# **Conclusion and Future Directions**



## **Chapter 6**

# **Sustainability Report**

### **6.1 Environmental**

### **6.2 Economic**

### **6.3 Social**

#### **6.3.1 Impacto Personal**

#### **6.3.2 Impacto Social**

#### **6.3.3 Riesgos Sociales**