UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC)
– BARCELONATECH

BACHELOR'S THESIS

Using Random Fourier Features with Random Forest

Author:
Albert RIBES

Supervisor:
Lluís A. BELANCHE

Computer Science

April 7, 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

List of Figures

Chapter 1

Introduction

1.1 Problem to solve

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

Machine Learning has shown it can be very useful when trying to predict a numerical or categorical variable based on some input data. It is able to define mathematical and statistical models which can help us on a lot of different fields, specially those that still need the presence of a human to take some decision.
However, in most of the situations a trade-off needs to be made between the ammount of precision in the predictions of the model (the accuracy) and the ammount of time the model needs to define the prediction function based on the data. There are many classification and regression problems that still require very powerful computers and a lot of “training time” in order to produce decent answers, and some of them still can’t be solved with a sufficient level of accuracy.
A lot of research is being done in the scientific community trying to improve this trade-off. New models are defined, variations to old ones, techniques to approximate more complex methods, etc. As the field progresses, it is possible to deal with problems that where out the scope of Machile Learning. But still, a lot of work needs to be done.

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

These days being able to learn from the data has many important applications. Big companies make use of Machine Learning techniques in order to be more efficient. Having good and cheap learning models helps them to have a better, faster and more efficient decision making.

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

The current development of Machine Learning has opened many fronts and techniques trying to solve many of the difficulties that the field has.
One known issue is the Bias-Variance dilemma. While solving a classification or regression problem, the expected generalization error of the models is the sum of three error terms: the bias, the variance and the irreductible error. While the last one, as the name sugests, can not be reduced, since it is caused by the inherent random noise in the data, the other two terms seem to have an inverse proportion: trying to reduce one of them increases the other one most of the times. In an attempt to reduce both of them (or at least their sum) there have been developed some ensemble methods. While they tend to show good results, their usage is mostly restricted to a small set of models. This is because they tend to work better on unstable models.
Another advance in Machine Learning has been the usage of kernel methods. They are useful to transform the data into another dimension with better properties, such as a clear dividing margin between classes of data. They are very effective, but their high computational costs has caused their use to be limited to just some specific problems or with a small number of instances. There are some less expensive approaches to approximate this methods, but they are not widely used.
And then there is a collection of classical algorithms which tend to have the advantage of being very simple and straightforward, although they don't usually get the highests scores.
There is a collection of techniques that have shown some good results by their own, but still they haven't been tested in combination with the others. If we could mix some of these methods

Chapter 2

Background Information and Theory

2.1 Machine Learning

- Una definición rápida
- 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

- No se basa en productos escalares
- Es extremadamente rápido
- Es más fácil de interpretar que otros modelos
- Es extremadamente inestable
- Cuando se hace un Random Forest, se randommiza un poco, de modo que árboles distintos entrenados con los mismos datos pueden ser destintos
- Es un modelo no lineal

2.2.2 Logistic Regression

Hola que tal

2.2.3 Support Vector Machines

- Inicialmente pensadas para clasificación en 2 clases
- Pero se puede más clases con *one-vs-rest* y también hay formas de hacer regresión
- Se basa únicamente en el producto escalar de sus entradas
- Intenta separar los datos con un híper-plano
- Actualmente es poco eficiente usarlas porque su coste s cúbico con la cantidad de entradas.
- Las fórmulas que quiere optimizar

2.3 Ensemble Methods

- Bagging
  - Inventado por Leo Breiman (referencia)
  - Pretende reducir el sesgo
  - Wikipedia dice que pretende reducir la varianza
  - Es el boosting el que pretende reducir el sesgo
  - Entrenamiento de los estimadores es independiente, se podría hacer en paralelo
  - Actualmente casi solo se usa con DT, debido a su inestabilidad
- Bootstrap
  - Intenta solucionar el problema de que para bagging es bueno que los estimadores sean distintos
  - Idealmente usaríamos un dataset distinto para cada estimador
  - Consiste en hacer un resalmpling con repetición
  - Si la cantidad de instancias del original es la misma que la de cada uno de los subconjuntos, se espera que la proporción de elementos úncos sea de $1 - \frac{1}{e} \approx 0.632$.
  - Si el conjunto original tiene n elementos, y tu haces un subconjunto de tamaño r, puedes esperar que la proporción de elementos del original que sí tienen presencia en el nuevo sea de $1 - e^{-\frac{r}{n}}$
- Random Forest

2.4 The kernel trick

- Teorema de Bochner
- El kernel RBF
  - Su fórmula es . . .
  - Equivalencia entre γ y σ
  - La noción de similitud que tiene
  - $\mathcal{H}$ es de dimensionalidad infinita
  - Permite ajustarse infinitamente a los datos, tuneando el híper-parámetro
  - σ más pequeño, más sobreajuste
  - γ más grande, más sobreajuste

2.5 Random Fourier Features

2.6 Nyström

Chapter 3

Project Development

3.1 General Idea

- Hemos visto que se puede sacar una aproximación aleatoria de la función implícita de un shift invariant kernel. Esto tiene 2 ventajas
  - Podemos transformar los datos directamente
  - Podemos producir pequeñas variaciones de un mismo dataset, todas ellas válidas
- Las 4 tipos de modelos que he definido. Referencia a la foto
- ¿Por qué he cogido estos 4 modelos? ¿No podrían haber sido otros? ¿Que tienen estos de bueno? Me he inspirado en Random Forest
- Hay por ahí algún paper que compara RFF y Nyström

3.1.1 State of the art con las RFF

- Se ha trabajado poco con ellas. Solo he encontrado 2 usos:
  - Stacked kernel network (referencia): usarlas junto a una red neuronal para tener más niveles de aprendizaje no lineal
  - RFF with SVM (referencia): usar una SVM sin kernel con los datos mapeados usando RFF

3.1.2 State of the art con las Nyström

3.2 Hyper-parameters

- Existen los siguientes:
  - min-impurity-decrease para DT
  - C para SVM
  - gamma para RFF y Nyström
  - cantidad de features para RFF y Nyström

- cantidad de estimadores para ensembles

- Hemos usado los siguientes valores:
  - Cantidad de features a 500
  - Cantidad de estimadores a 50
  - En modelos simples, el parámetro por crossvalidation
  - En modelos simples con RFF, el parámetro por crossvalidation y una gamma que sobreajuste
  - En modelos con ensemble, parámetros que sobreajusten y la gamma por crossvalidation
  - En RBF-SVM, la gamma por gamest y el parámetro por crossvalidation

3.3 Hypothesis

1. Podemos aproximar bien una RBF-SVM
2. Puede tener sentido hacer ensembles con otros modelos a DT
3. RFF + Bootstrap puede ser malo
4. Si el modelo no se basa en productos escalares no se feneficiará tanto

3.3.1 Planteamiento de los experimentos

1. Hipótesis: Aproximar RBF-SVM
   1.1. Comparar una RBF-SVM con SVM normal que use RFF
2. Hipótesis: Ensembles con otros
   2.1. Logit normal vs. Logit con RFF
   2.2. Logit normal vs. Logit con RFF Black Bag
   2.3. Logit normal vs. Logit con RFF Grey Bag
   2.4. Logit normal vs. Logit con RFF Grey Ensemble
   2.5. Linear-SVM vs Linear-SVM con RFF
   2.6. Linear-SVM vs Linear-SVM con RFF Black Bag
   2.7. Linear-SVM vs Linear-SVM con RFF Grey Bag
   2.8. Linear-SVM vs Linear-SVM con RFF Grey Ensemble
3. Hipótesis: RFF + Bootstrap
   3.1. Logit con RFF Grey Bag vs Logit con RFF Grey Ensemble
   3.2. Logit con RFF Black Bag vs Logit con RFF Black Ensemble (los dos con un solo estimador)

3.3. Linear-SVM con RFF Grey Bag vs Linear-SVM con RFF Grey Ensemble
3.4. Linear-SVM con RFF Black Bag vs Linear-SVM con RFF Black Ensemble (los

4. Hipótesis: DT + RFF
    - DT vs DT con RFF
    - DT vs DT con RFF Black Bag
    - DT vs DT con RFF Black Ensemble
    - DT vs DT con RFF Grey Bag
    - DT vs DT con RFF Grey Ensemble

3.4 Datasets

- 8 Datasets
- Normalizados
- Únicamente tienen variables numéricas, no categóricas
- Únicamente problemas de clasificación
- Algunas cosas particulares que he hecho:
    - Mezclar datos de train y de test para luego hacer mi propia separación
    - Cuando había poca presencia de una clase, hacer un resampling para igualar las cantidades
    - No trabajar cosas como el skiwness o los outliers
    - Eliminar columnas en las que todo eran 0
    - Reducir el conjunto de instancias

Chapter 4

Experimental Results

4.1 Enfrentar resultados 2 a 2

4.2 Contrastar hipótesis con resultados

Chapter 5

Conclusion and Future Directions

- Problemas de regresión
- Aproximar otros kernels a RBF
- Ver el comportamiento con problemas que no sean tan bonitos (con missings, clases desbalanceadas, etc)
- Otros tipos de ensembles, como el boosting

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

Appendix A

Results of experiment 1.1

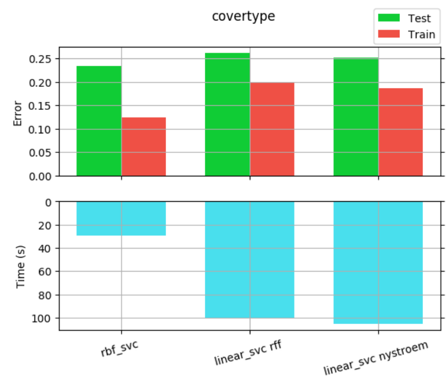

FIGURE A.1: RBF-SVM and Linear-SVM with RFF and Nyström

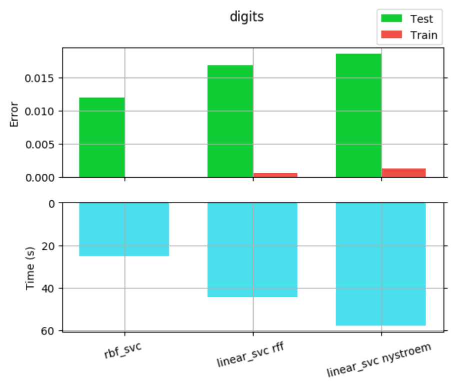

FIGURE A.2: RBF-SVM and Linear-SVM with RFF and Nyström

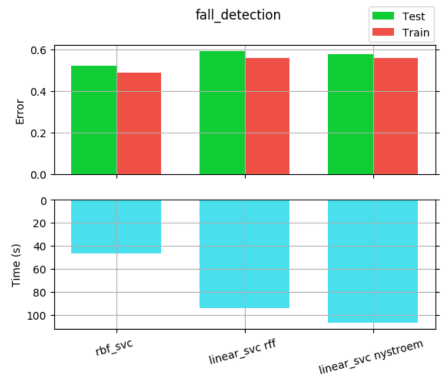

FIGURE A.3: RBF-SVM and Linear-SVM with RFF and Nyström

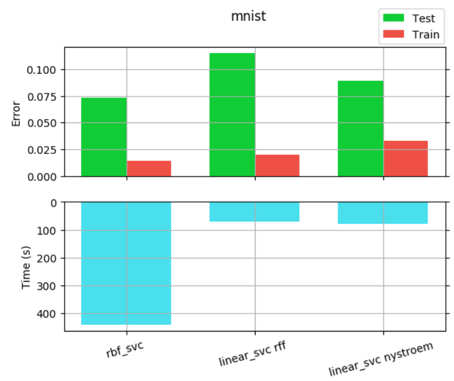

FIGURE A.4: RBF-SVM and Linear-SVM with RFF and Nyström

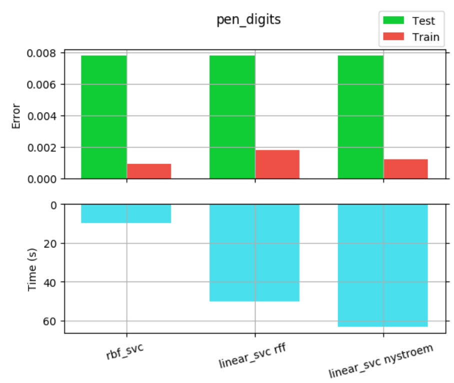

FIGURE A.5: RBF-SVM and Linear-SVM with RFF and Nyström

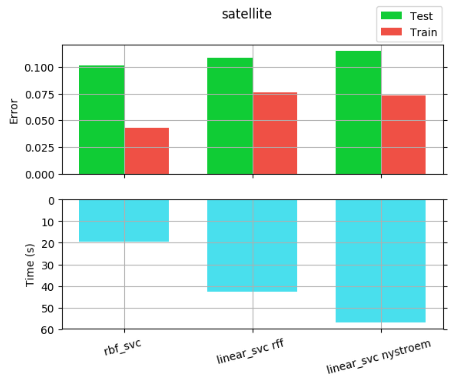

FIGURE A.6: RBF-SVM and Linear-SVM with RFF and Nyström

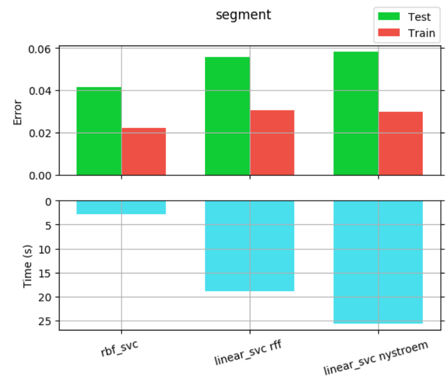

FIGURE A.7: RBF-SVM and Linear-SVM with RFF and Nyström

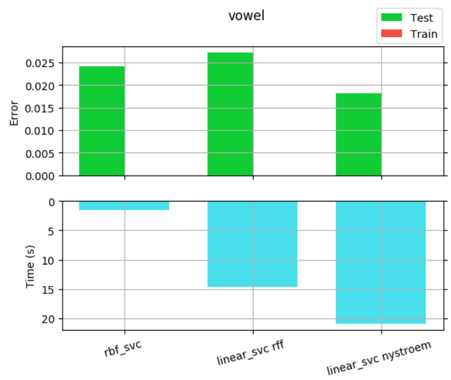

FIGURE A.8: RBF-SVM and Linear-SVM with RFF and Nyström

Appendix B

Results of experiment 2.1

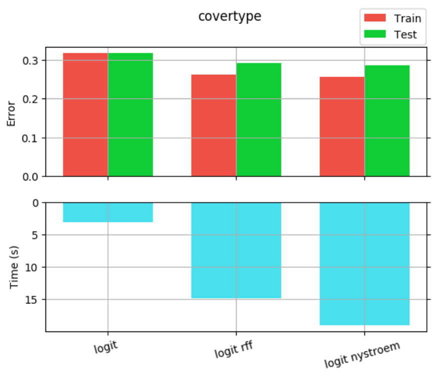

FIGURE B.1: Normal Logistic Regression and witdh RFF and Nyström

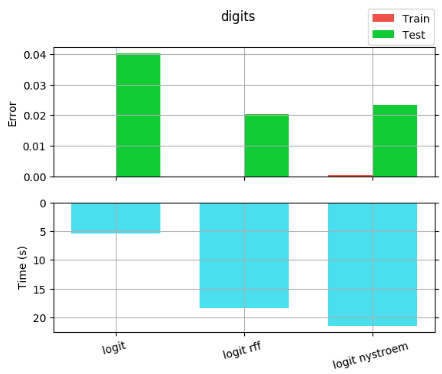

FIGURE B.2: Normal Logistic Regression and witdh RFF and Nyström

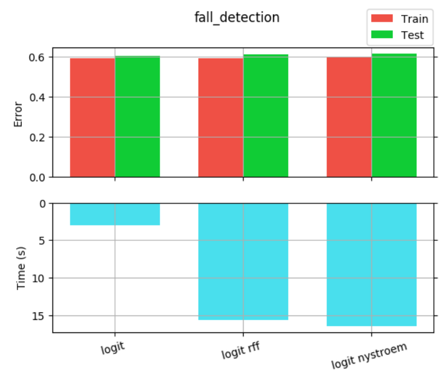

FIGURE B.3: Normal Logistic Regression and witdh RFF and Nyström

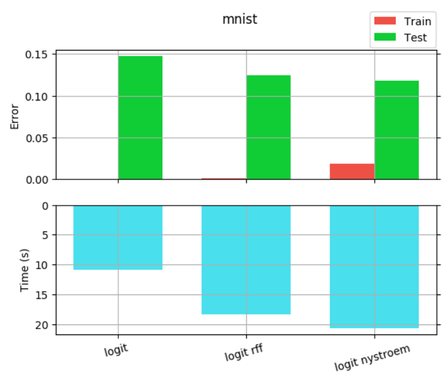

FIGURE B.4: Normal Logistic Regression and witdh RFF and Nyström

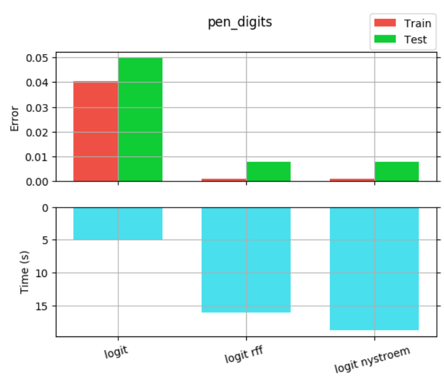

FIGURE B.5: Normal Logistic Regression and witdh RFF and Nyström

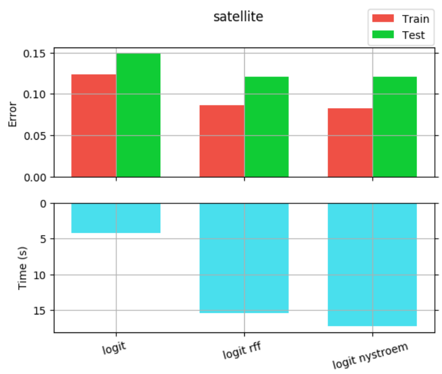

FIGURE B.6: Normal Logistic Regression and witdh RFF and Nyström

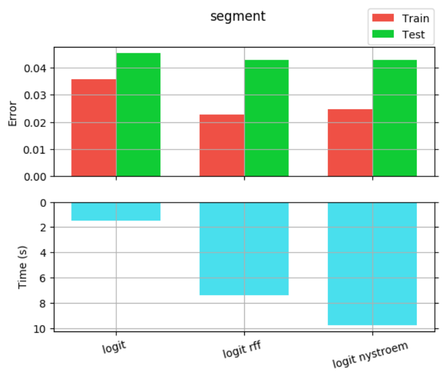

FIGURE B.7: Normal Logistic Regression and witdh RFF and Nyström

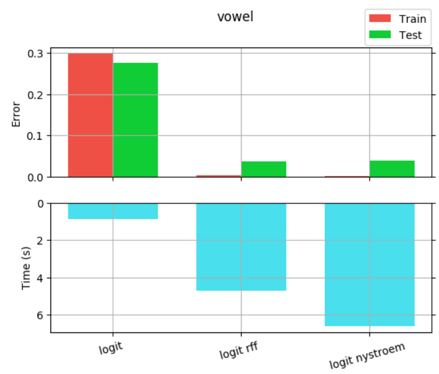

FIGURE B.8: Normal Logistic Regression and witdh RFF and Nyström

Appendix C

Results of experiment 2.2

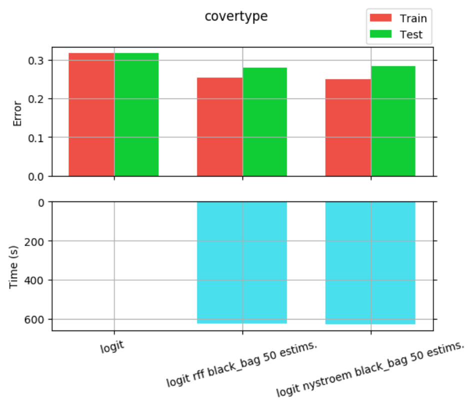

FIGURE C.1: Logistic Regression with Black Bag model

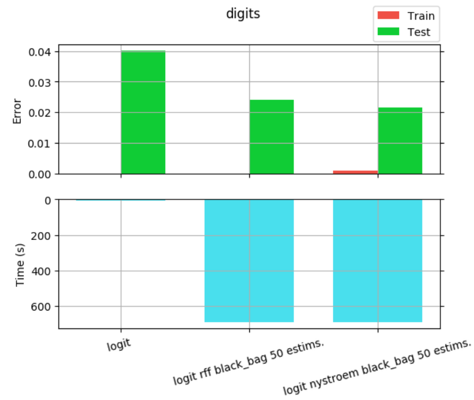

FIGURE C.2: Logistic Regression with Black Bag model

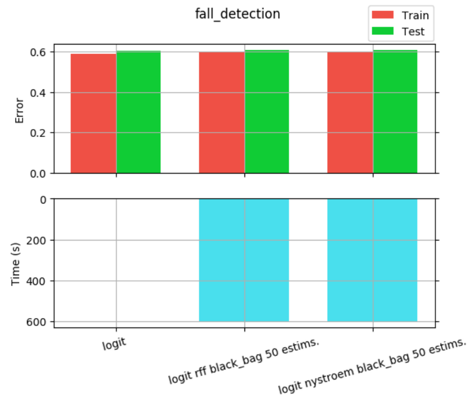

FIGURE C.3: Logistic Regression with Black Bag model

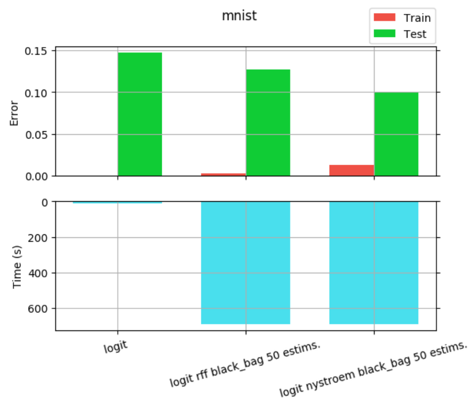

FIGURE C.4: Logistic Regression with Black Bag model

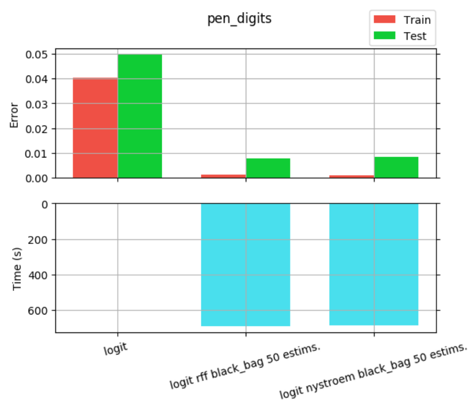

FIGURE C.5: Logistic Regression with Black Bag model

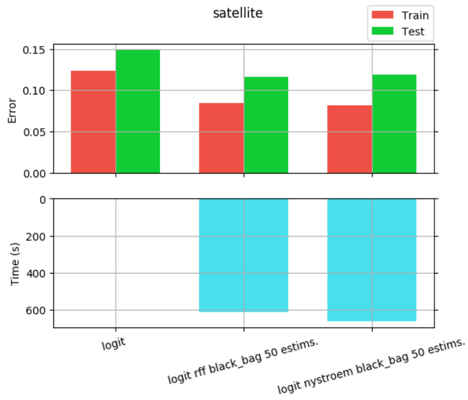

FIGURE C.6: Logistic Regression with Black Bag model

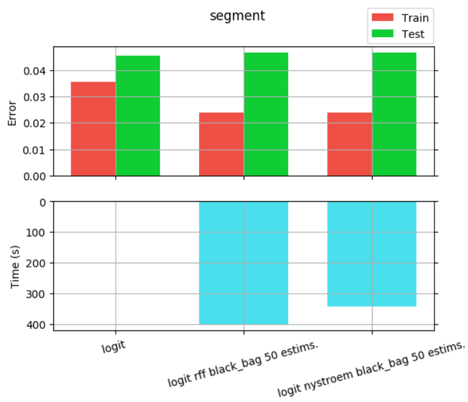

FIGURE C.7: Logistic Regression with Black Bag model

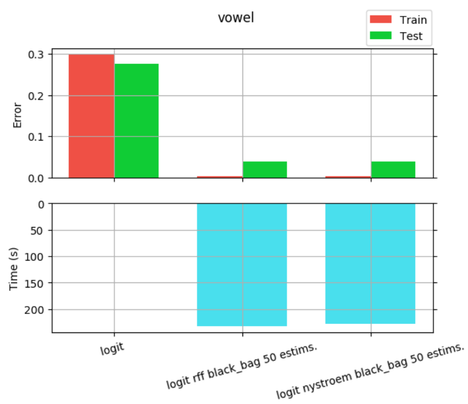

FIGURE C.8: Logistic Regression with Black Bag model

Appendix D

Results of experiment 2.3

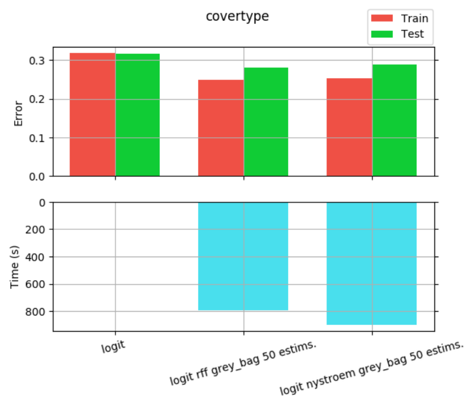

FIGURE D.1: Logistic Regression with Grey Bag model

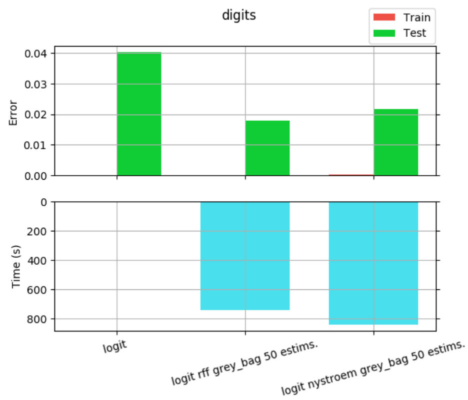

FIGURE D.2: Logistic Regression with Grey Bag model

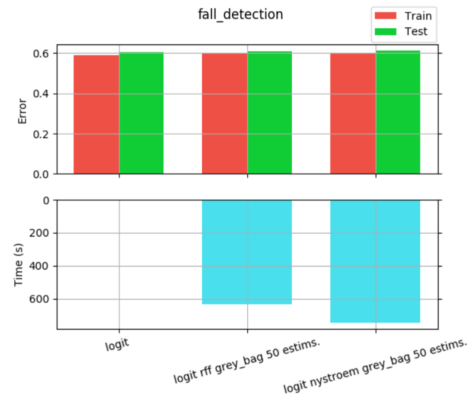

FIGURE D.3: Logistic Regression with Grey Bag model

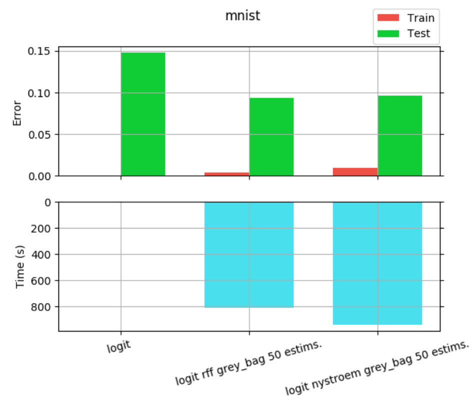

FIGURE D.4: Logistic Regression with Grey Bag model

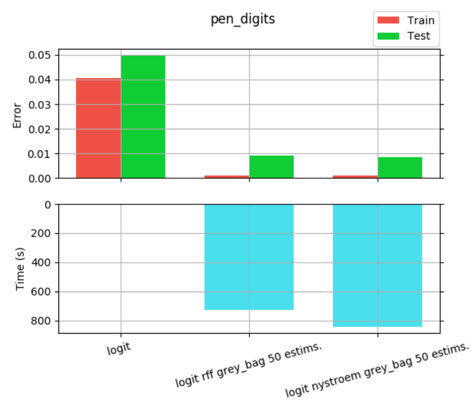

FIGURE D.5: Logistic Regression with Grey Bag model

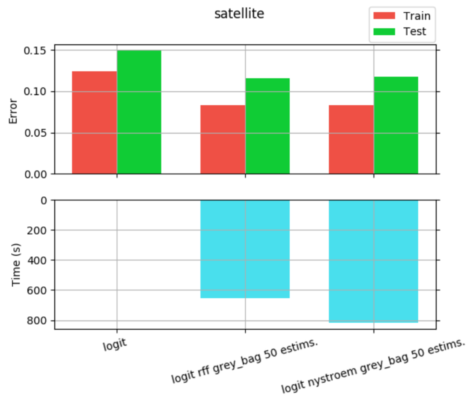

FIGURE D.6: Logistic Regression with Grey Bag model

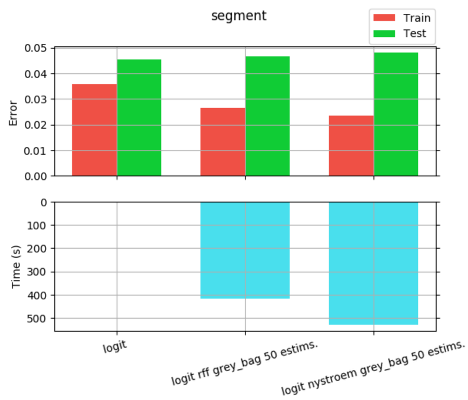

FIGURE D.7: Logistic Regression with Grey Bag model

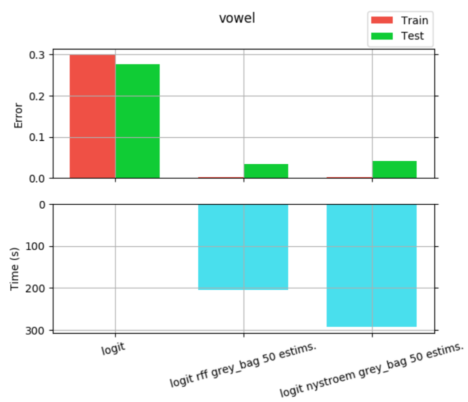

FIGURE D.8: Logistic Regression with Grey Bag model

Appendix E

Results of experiment 2.4

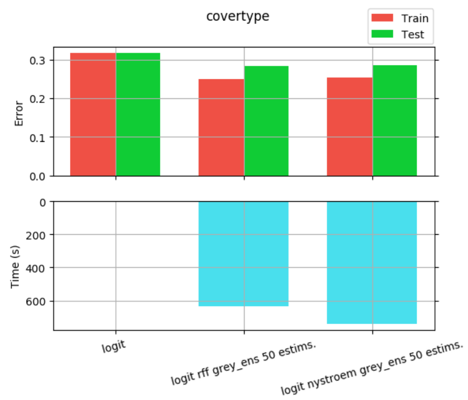

FIGURE E.1: Logistic Regression with Grey Ensemble model

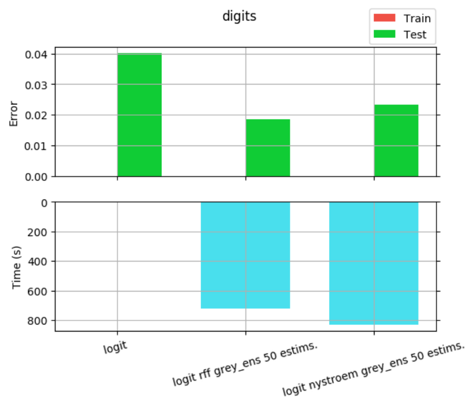

FIGURE E.2: Logistic Regression with Grey Ensemble model

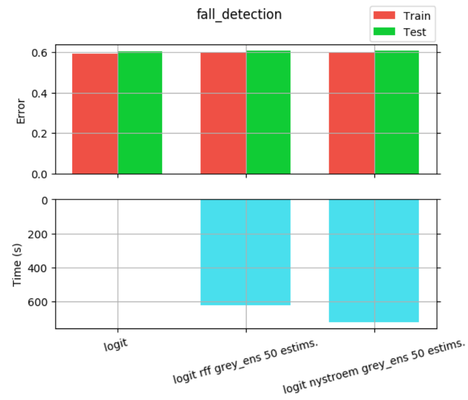

FIGURE E.3: Logistic Regression with Grey Ensemble model

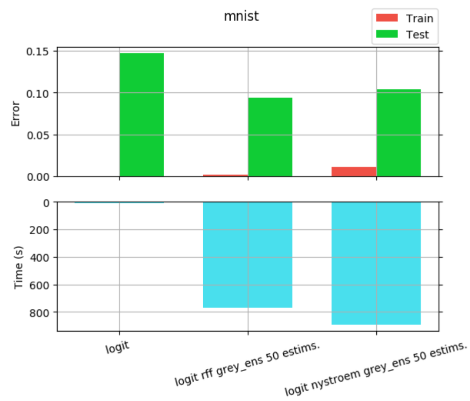

FIGURE E.4: Logistic Regression with Grey Ensemble model

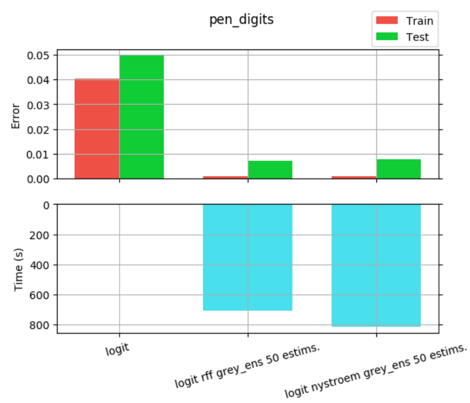

FIGURE E.5: Logistic Regression with Grey Ensemble model

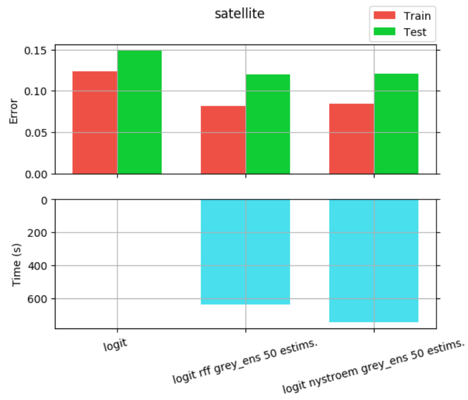

FIGURE E.6: Logistic Regression with Grey Ensemble model

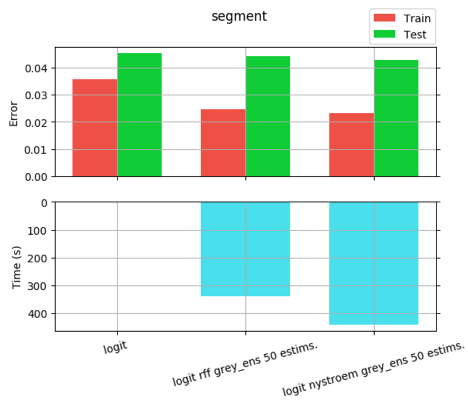

FIGURE E.7: Logistic Regression with Grey Ensemble model

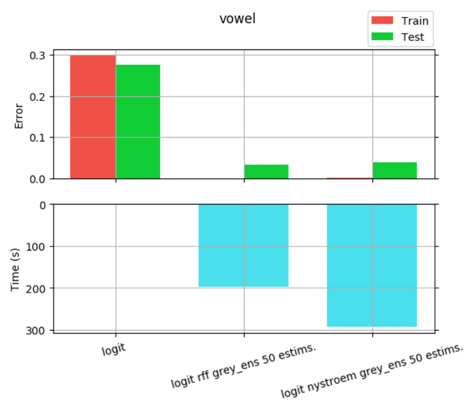

FIGURE E.8: Logistic Regression with Grey Ensemble model

Appendix F

Results of experiment 2.5

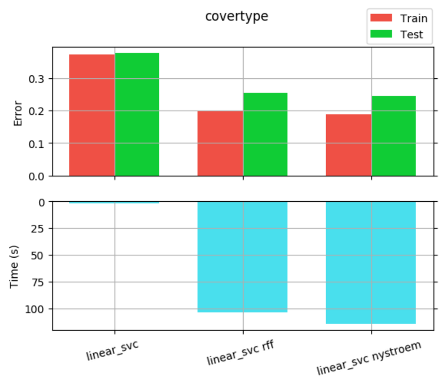

FIGURE F.1: Linear-SVC with RFF and Nyström

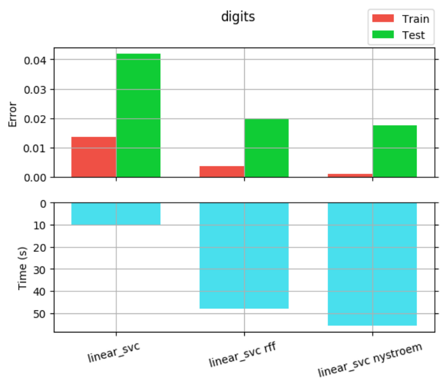

FIGURE F.2: Linear-SVC with RFF and Nyström

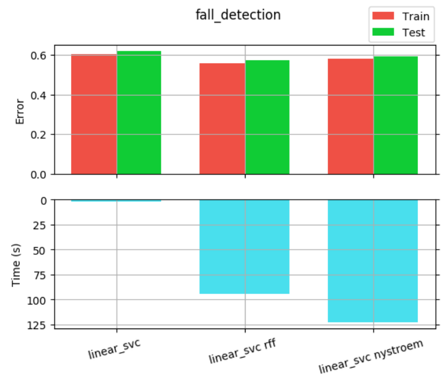

FIGURE F.3: Linear-SVC with RFF and Nyström

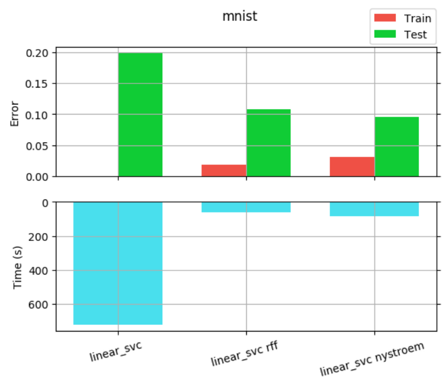

FIGURE F.4: Linear-SVC with RFF and Nyström

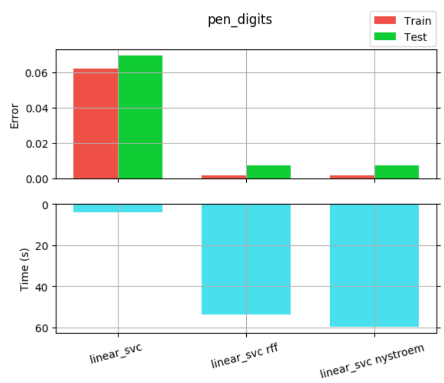

FIGURE F.5: Linear-SVC with RFF and Nyström

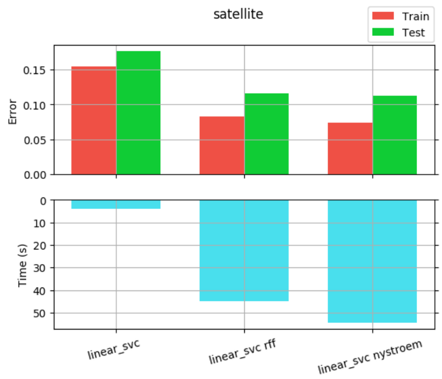

FIGURE F.6: Linear-SVC with RFF and Nyström

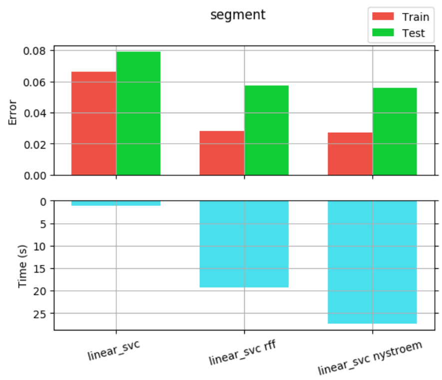

FIGURE F.7: Linear-SVC with RFF and Nyström

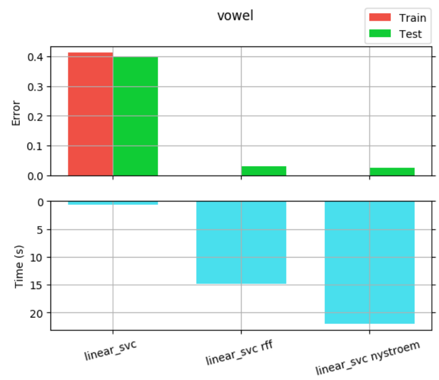

FIGURE F.8: Linear-SVC with RFF and Nyström

Appendix G

Results of experiment 2.6

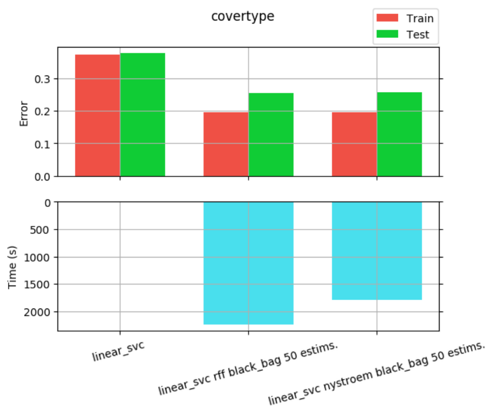

FIGURE G.1: Linear-SVM with Black Bag model

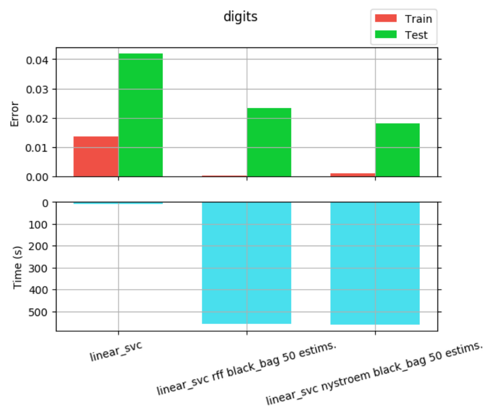

FIGURE G.2: Linear-SVM with Black Bag model

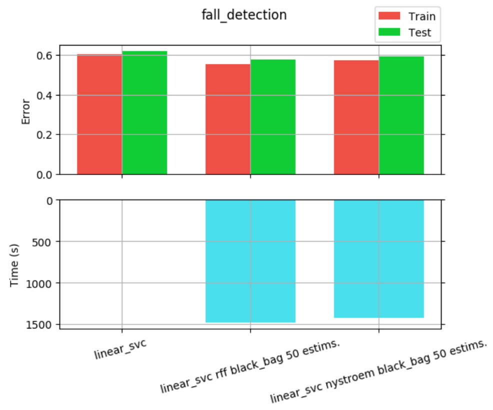

FIGURE G.3: Linear-SVM with Black Bag model

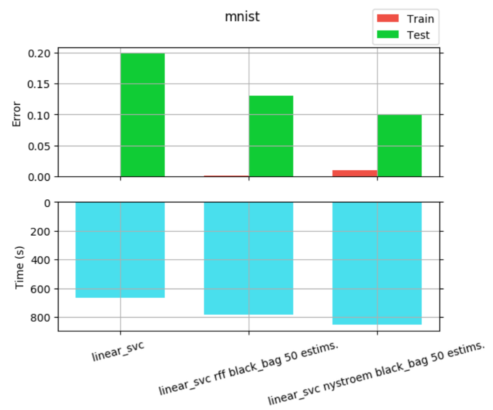

FIGURE G.4: Linear-SVM with Black Bag model

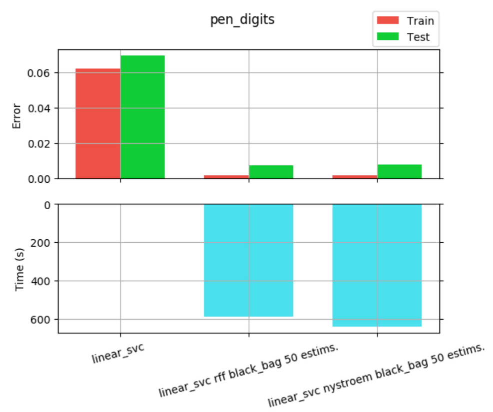

FIGURE G.5: Linear-SVM with Black Bag model

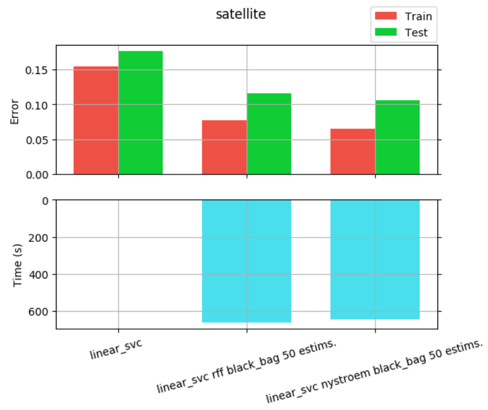

FIGURE G.6: Linear-SVM with Black Bag model

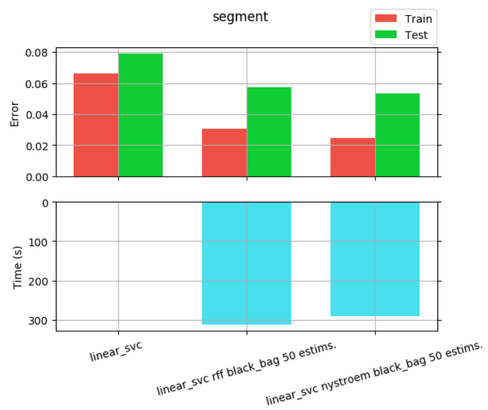

FIGURE G.7: Linear-SVM with Black Bag model

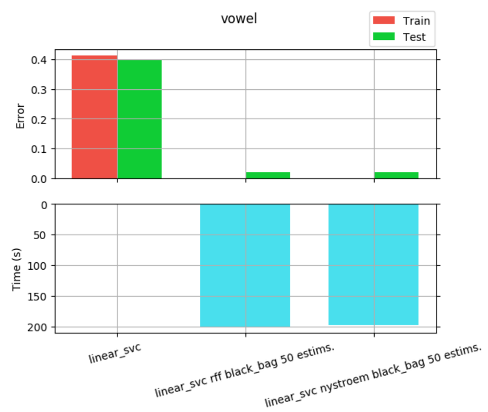

FIGURE G.8: Linear-SVM with Black Bag model

Appendix H

Results of experiment 2.2

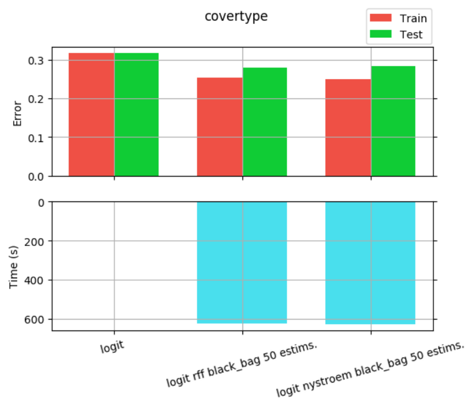

FIGURE H.1: Logistic Regression with Black Bag model

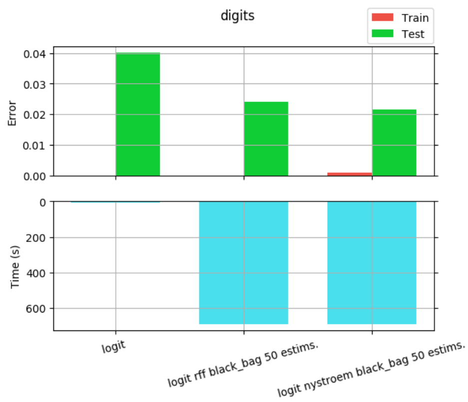

FIGURE H.2: Logistic Regression with Black Bag model

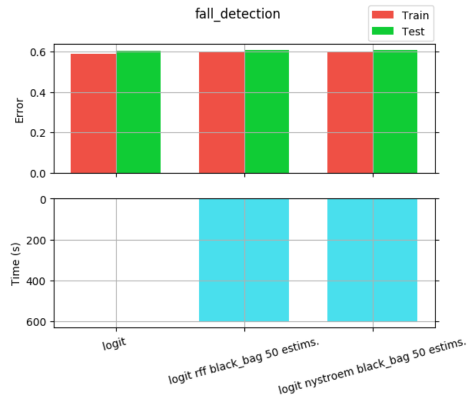

FIGURE H.3: Logistic Regression with Black Bag model

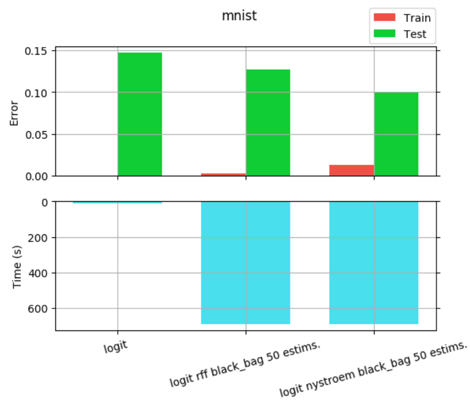

FIGURE H.4: Logistic Regression with Black Bag model

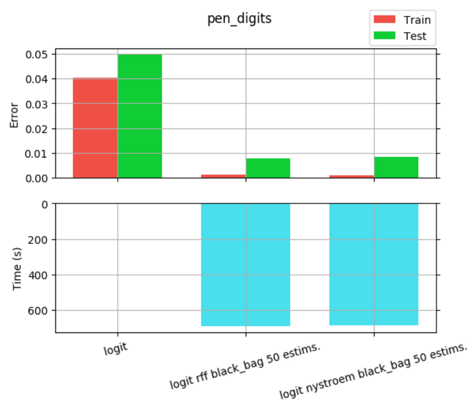

FIGURE H.5: Logistic Regression with Black Bag model

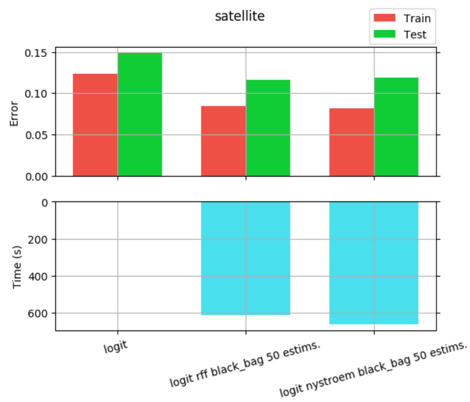

FIGURE H.6: Logistic Regression with Black Bag model

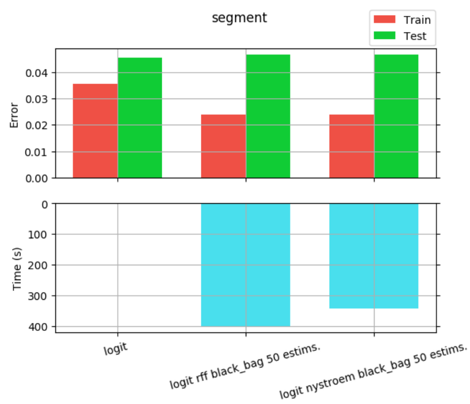

FIGURE H.7: Logistic Regression with Black Bag model

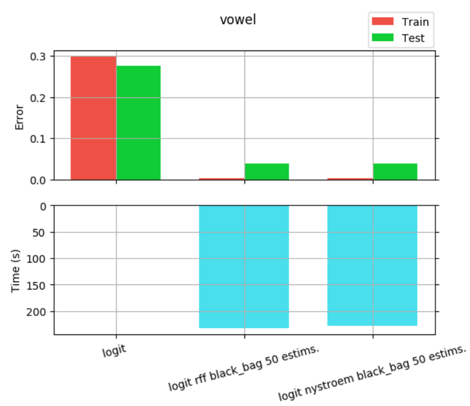

FIGURE H.8: Logistic Regression with Black Bag model

Appendix I

Results of experiment 2.2

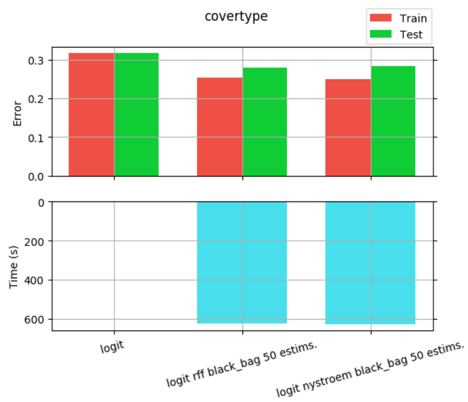

FIGURE I.1: Logistic Regression with Black Bag model

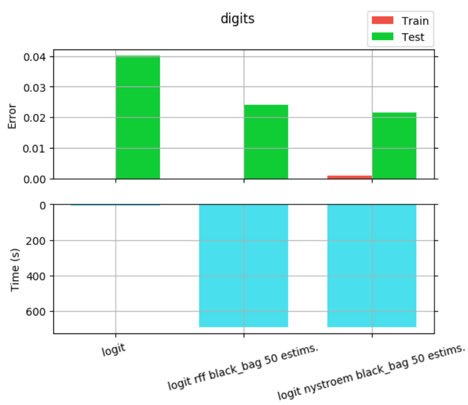

FIGURE I.2: Logistic Regression with Black Bag model

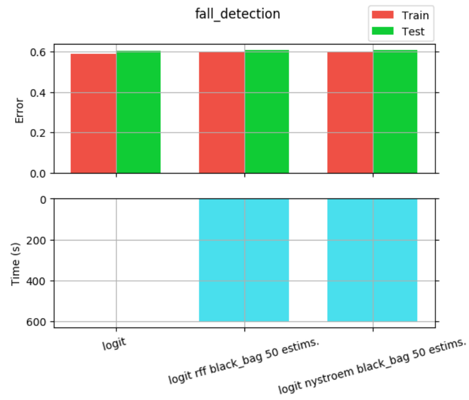

FIGURE I.3: Logistic Regression with Black Bag model

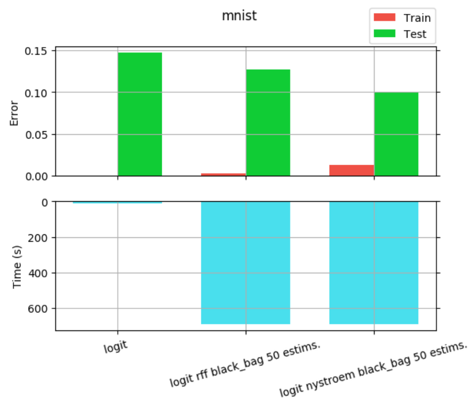

FIGURE I.4: Logistic Regression with Black Bag model

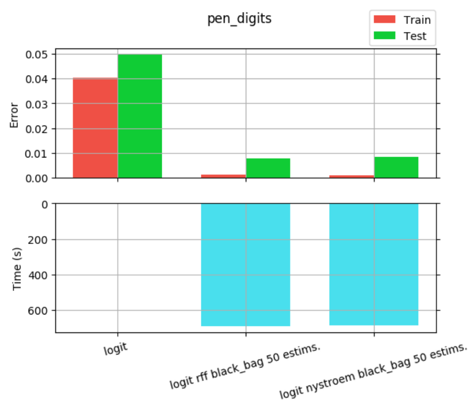

FIGURE I.5: Logistic Regression with Black Bag model

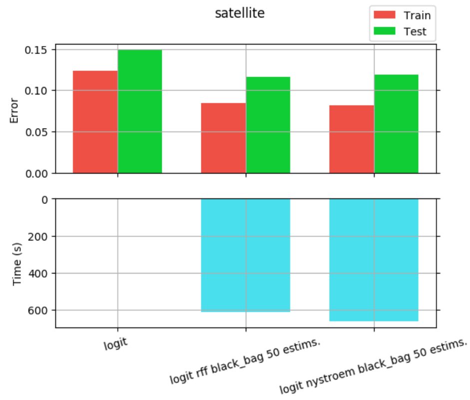

FIGURE I.6: Logistic Regression with Black Bag model

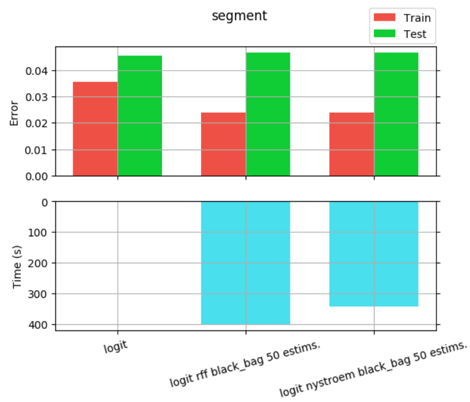

FIGURE I.7: Logistic Regression with Black Bag model

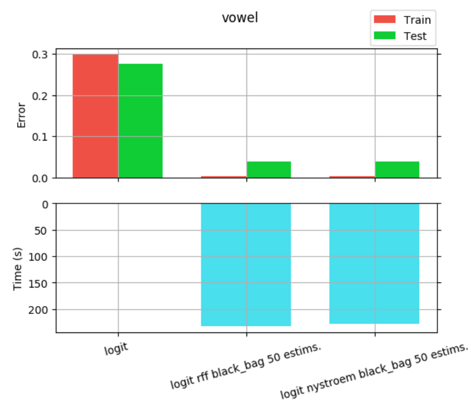

FIGURE I.8: Logistic Regression with Black Bag model