

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

April 3, 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

Abstract	iii
1 Introduction	1
1.1 Problem to solve	1
1.2 Why is it important?	1
1.3 Project proposal	1
2 Background Information and Theory	3
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	4
2.3 Ensemble Methods	4
2.4 The kernel trick	4
2.5 Random Fourier Features	5
2.6 Nyström	5
3 Project Development	7
3.1 General Idea	7
3.1.1 State of the art con las RFF	7
3.1.2 State of the art con las Nyström	7
3.2 Hyper-parameters	7
3.3 Hypothesis	8
3.3.1 Planteamiento de los experimentos	8
3.4 Datasets	9
4 Experimental Results	11
4.1 Enfrentar resultados 2 a 2	11
4.2 Contrastar hipótesis con resultados	11
5 Conclusion and Future Directions	13
6 Sustainability Report	15
6.1 Environmental	15
6.2 Economic	15
6.3 Social	15
6.3.1 Impacto Personal	15
6.3.2 Impacto Social	15
6.3.3 Riesgos Sociales	15
A Results of experiment 1.1	17
B Results of experiment 2.1	21

C	Results of experiment 2.2	25
D	Results of experiment 2.3	29
E	Results of experiment 2.4	33
F	Results of experiment 2.5	37
G	Results of experiment 2.6	41
H	Results of experiment 2.2	45
I	Results of experiment 2.2	49

List of Figures

A.1	1.1 coertype	17
A.2	1.1 digits	17
A.3	1.1 fall_detection	18
A.4	1.1 mnist	18
A.5	1.1 pen_digits	18
A.6	1.1 satellite	19
A.7	1.1 segment	19
A.8	1.1 vowel	19
B.1	2.1 coertype	21
B.2	2.1 digits	21
B.3	2.1 fall_detection	22
B.4	2.1 mnist	22
B.5	2.1 pen_digits	22
B.6	2.1 satellite	23
B.7	2.1 segment	23
B.8	2.1 vowel	23
C.1	2.2 coertype	25
C.2	2.2 digits	25
C.3	2.2 fall_detection	26
C.4	2.2 mnist	26
C.5	2.2 pen_digits	26
C.6	2.2 satellite	27
C.7	2.2 segment	27
C.8	2.2 vowel	27
D.1	2.3 coertype	29
D.2	2.3 digits	29
D.3	2.3 fall_detection	30
D.4	2.3 mnist	30
D.5	2.3 pen_digits	30
D.6	2.3 satellite	31
D.7	2.3 segment	31
D.8	2.3 vowel	31
E.1	2.4 coertype	33
E.2	2.4 digits	33
E.3	2.4 fall_detection	34
E.4	2.4 mnist	34
E.5	2.4 pen_digits	34
E.6	2.4 satellite	35
E.7	2.4 segment	35

E.8	2.4 vowel	35
F.1	2.5 coertype	37
F.2	2.5 digits	37
F.3	2.5 fall_detection	38
F.4	2.5 mnist	38
F.5	2.5 pen_digits	38
F.6	2.5 satellite	39
F.7	2.5 segment	39
F.8	2.5 vowel	39
G.1	2.6 coertype	41
G.2	2.6 digits	41
G.3	2.6 fall_detection	42
G.4	2.6 mnist	42
G.5	2.6 pen_digits	42
G.6	2.6 satellite	43
G.7	2.6 segment	43
G.8	2.6 vowel	43
H.1	2.2 coertype	45
H.2	2.2 digits	45
H.3	2.2 fall_detection	46
H.4	2.2 mnist	46
H.5	2.2 pen_digits	46
H.6	2.2 satellite	47
H.7	2.2 segment	47
H.8	2.2 vowel	47
I.1	2.2 coertype	49
I.2	2.2 digits	49
I.3	2.2 fall_detection	50
I.4	2.2 mnist	50
I.5	2.2 pen_digits	50
I.6	2.2 satellite	51
I.7	2.2 segment	51
I.8	2.2 vowel	51

List of Tables

List of Abbreviations

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

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

- 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 distintos
- 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 hiper-plano
- Actualmente es poco eficiente usarlas porque su coste es 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
 - 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 resampling 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 únicos 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 hiperparámetro
- 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

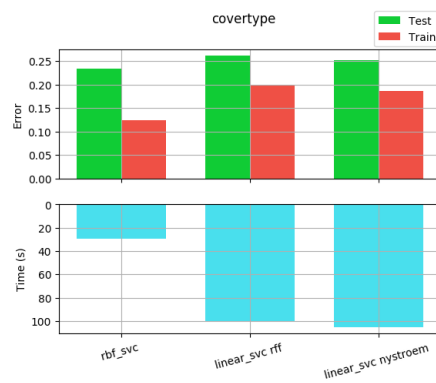


FIGURE A.1: RBF-SVM and Linear-SVM with RFF and Nyström

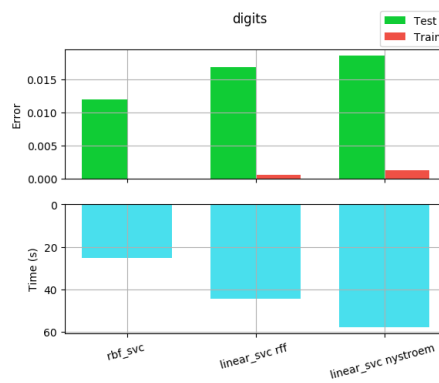


FIGURE A.2: RBF-SVM and Linear-SVM with RFF and Nyström

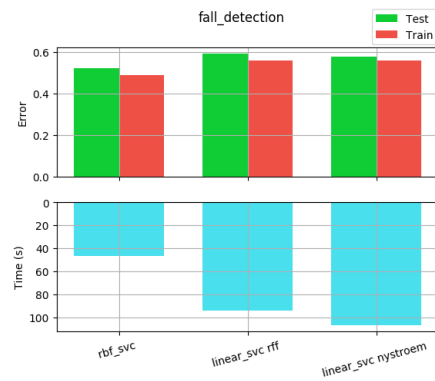


FIGURE A.3: RBF-SVM and Linear-SVM with RFF and Nyström

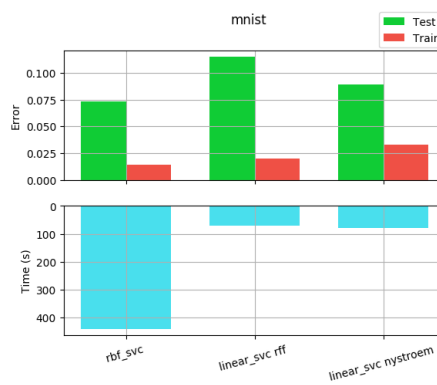


FIGURE A.4: RBF-SVM and Linear-SVM with RFF and Nyström

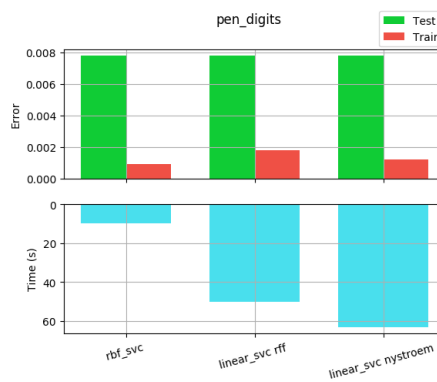


FIGURE A.5: RBF-SVM and Linear-SVM with RFF and Nyström

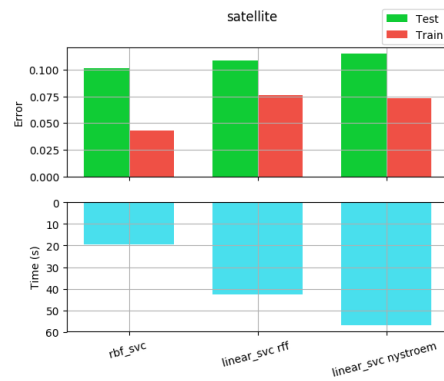


FIGURE A.6: RBF-SVM and Linear-SVM with RFF and Nyström

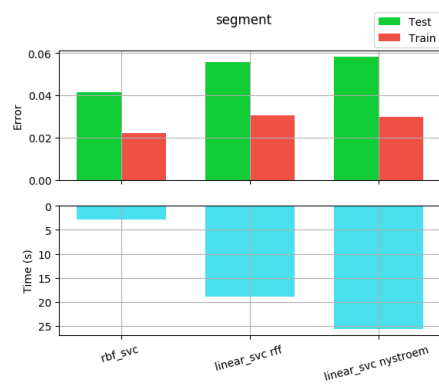


FIGURE A.7: RBF-SVM and Linear-SVM with RFF and Nyström

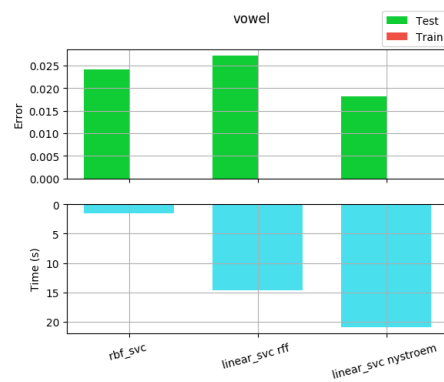


FIGURE A.8: RBF-SVM and Linear-SVM with RFF and Nyström

Appendix B

Results of experiment 2.1

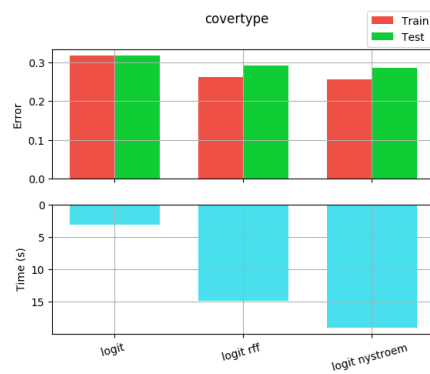


FIGURE B.1: Normal Logistic Regression and with RFF and Nyström

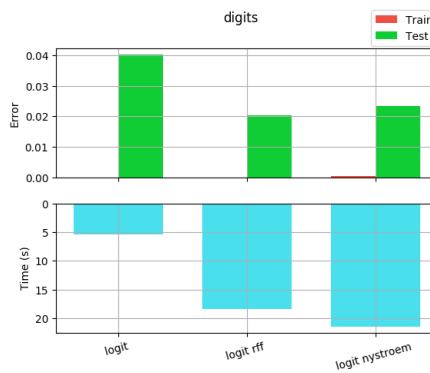


FIGURE B.2: Normal Logistic Regression and with RFF and Nyström

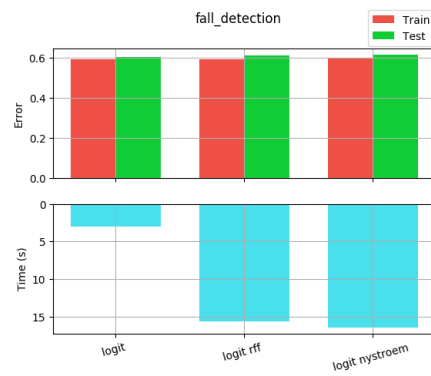


FIGURE B.3: Normal Logistic Regression and with RFF and Nyström

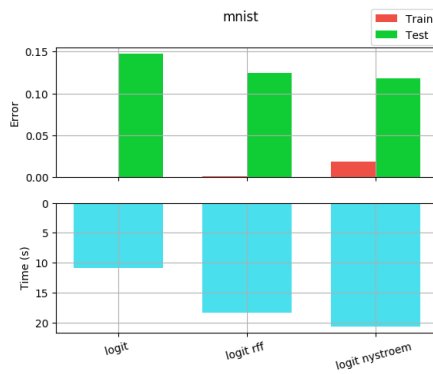


FIGURE B.4: Normal Logistic Regression and with RFF and Nyström

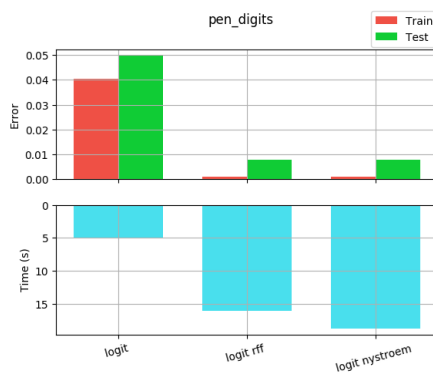


FIGURE B.5: Normal Logistic Regression and with RFF and Nyström

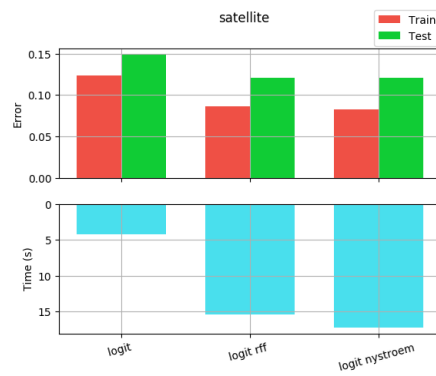


FIGURE B.6: Normal Logistic Regression and with RFF and Nyström

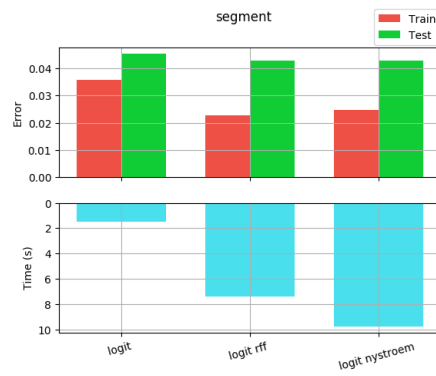


FIGURE B.7: Normal Logistic Regression and with RFF and Nyström

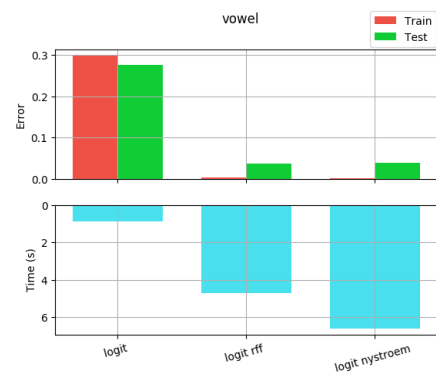


FIGURE B.8: Normal Logistic Regression and with RFF and Nyström

Appendix C

Results of experiment 2.2

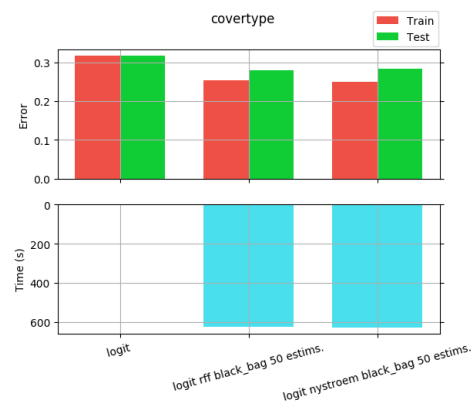


FIGURE C.1: Logistic Regression with Black Bag model

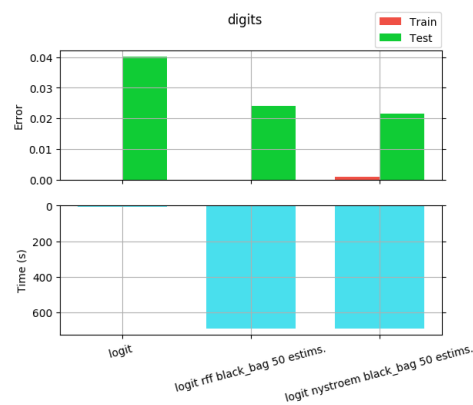


FIGURE C.2: Logistic Regression with Black Bag model

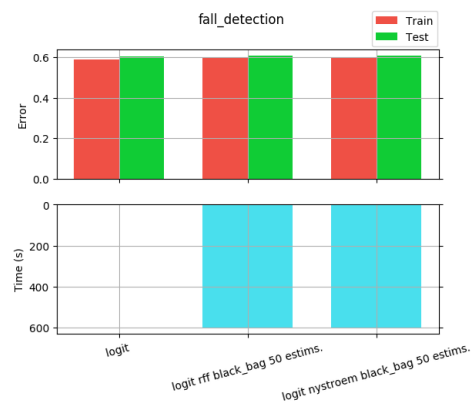


FIGURE C.3: Logistic Regression with Black Bag model

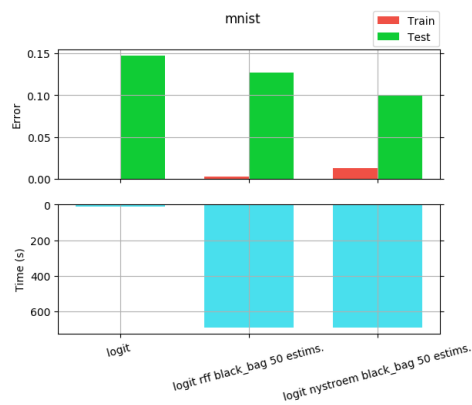


FIGURE C.4: Logistic Regression with Black Bag model

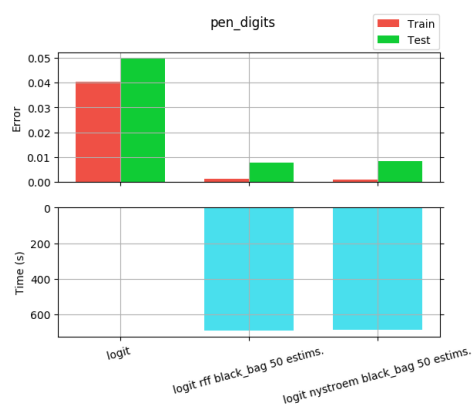


FIGURE C.5: Logistic Regression with Black Bag model

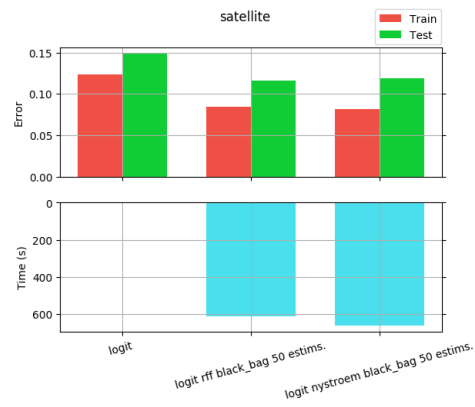


FIGURE C.6: Logistic Regression with Black Bag model

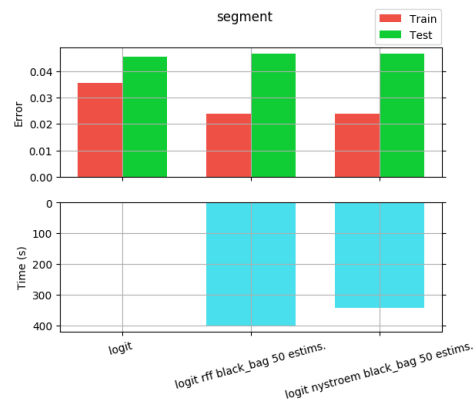


FIGURE C.7: Logistic Regression with Black Bag model

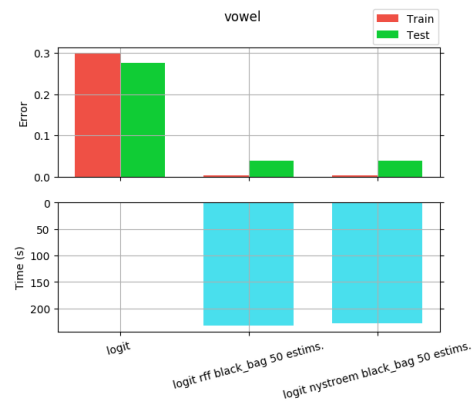


FIGURE C.8: Logistic Regression with Black Bag model

Appendix D

Results of experiment 2.3

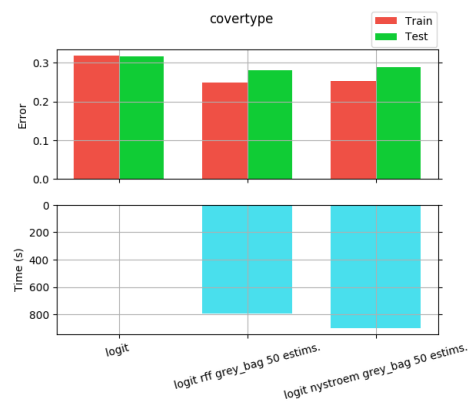


FIGURE D.1: Logistic Regression with Grey Bag model

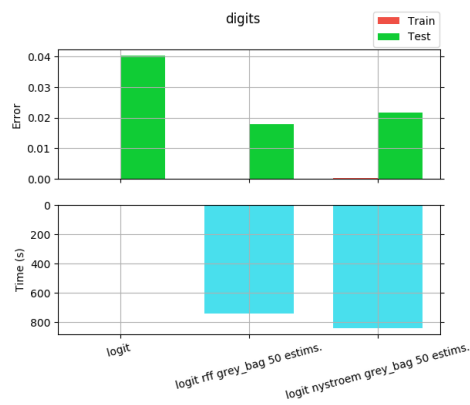


FIGURE D.2: Logistic Regression with Grey Bag model

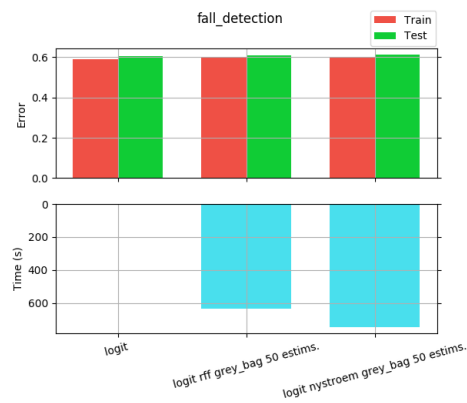


FIGURE D.3: Logistic Regression with Grey Bag model

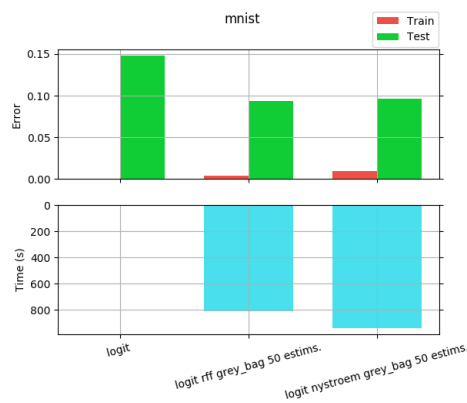


FIGURE D.4: Logistic Regression with Grey Bag model

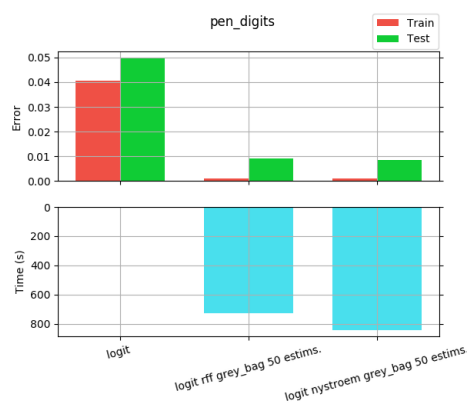


FIGURE D.5: Logistic Regression with Grey Bag model

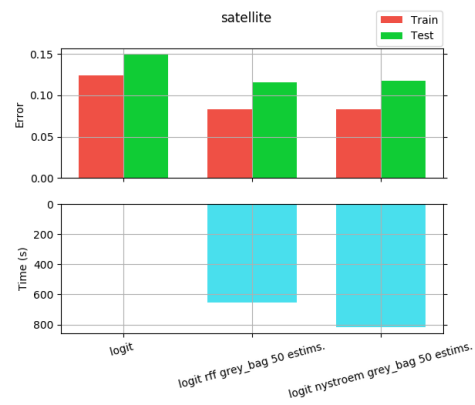


FIGURE D.6: Logistic Regression with Grey Bag model

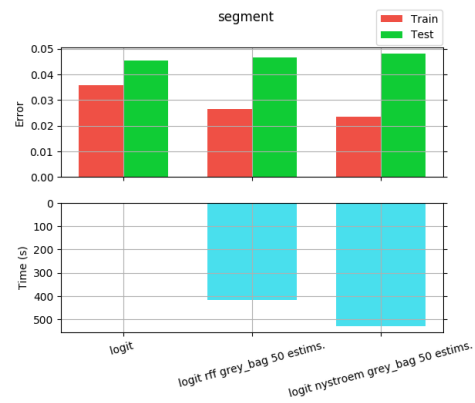


FIGURE D.7: Logistic Regression with Grey Bag model

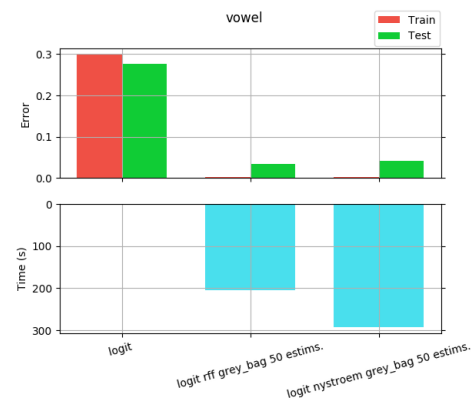


FIGURE D.8: Logistic Regression with Grey Bag model

Appendix E

Results of experiment 2.4

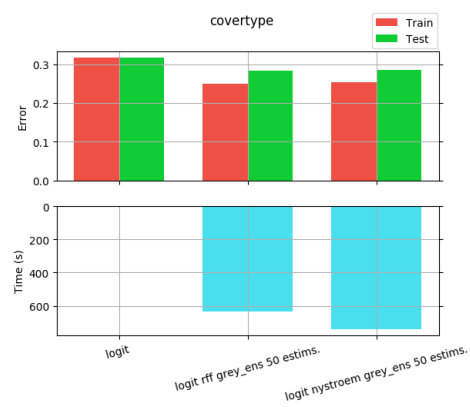


FIGURE E.1: Logistic Regression with Grey Ensemble model

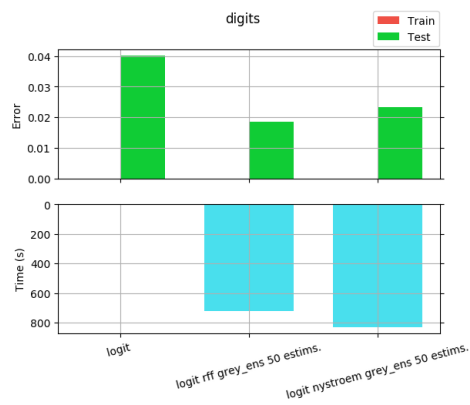


FIGURE E.2: Logistic Regression with Grey Ensemble model

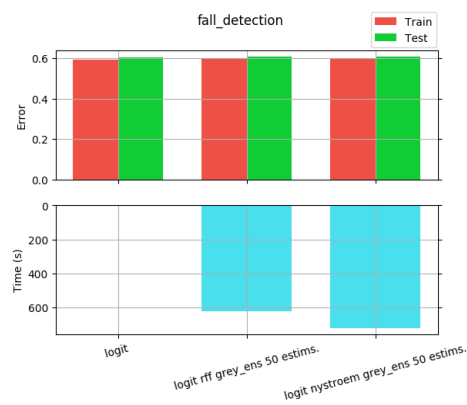


FIGURE E.3: Logistic Regression with Grey Ensemble model

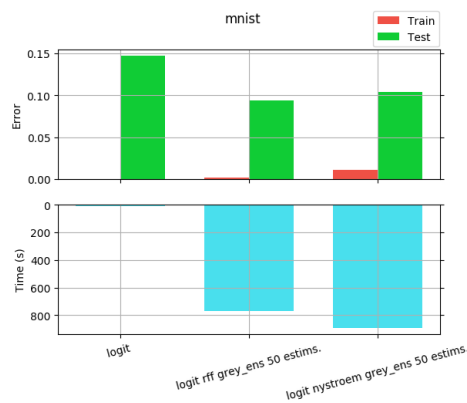


FIGURE E.4: Logistic Regression with Grey Ensemble model

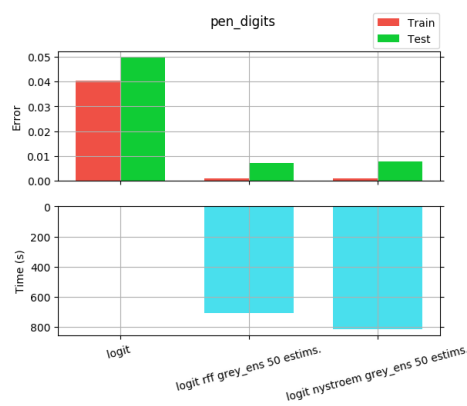


FIGURE E.5: Logistic Regression with Grey Ensemble model

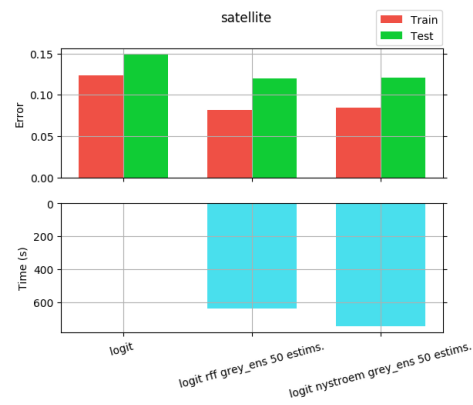


FIGURE E.6: Logistic Regression with Grey Ensemble model

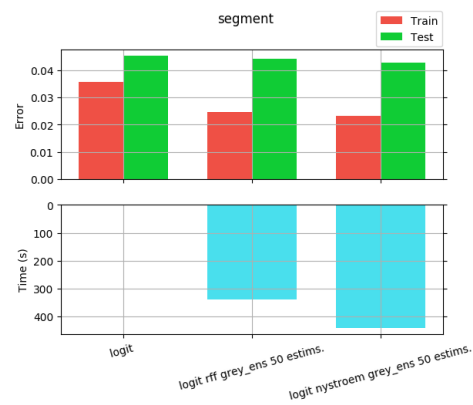


FIGURE E.7: Logistic Regression with Grey Ensemble model

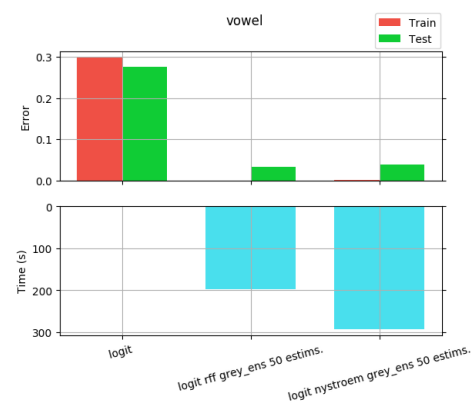


FIGURE E.8: Logistic Regression with Grey Ensemble model

Appendix F

Results of experiment 2.5

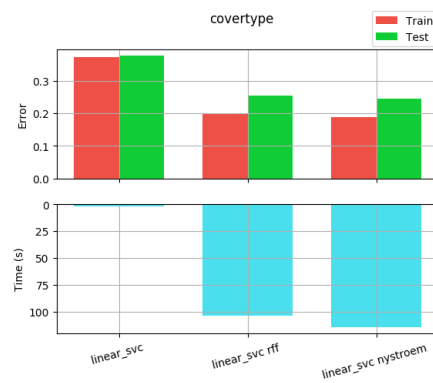


FIGURE F.1: Linear-SVC with RFF and Nyström

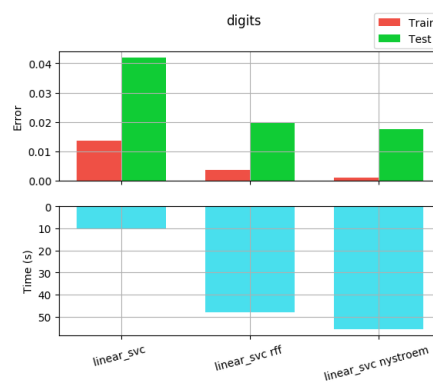


FIGURE F.2: Linear-SVC with RFF and Nyström

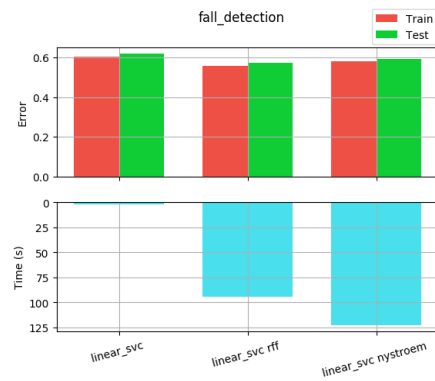


FIGURE F.3: Linear-SVC with RFF and Nyström

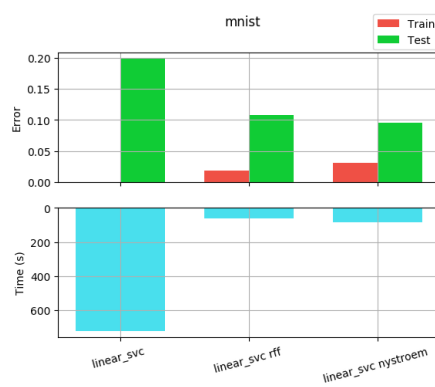


FIGURE F.4: Linear-SVC with RFF and Nyström

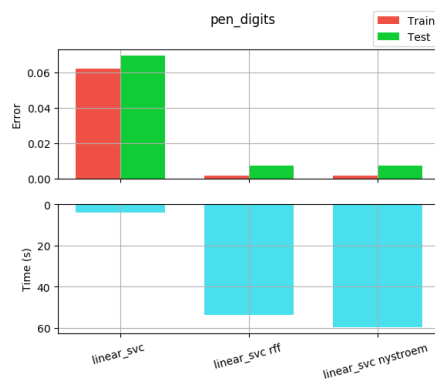


FIGURE F.5: Linear-SVC with RFF and Nyström

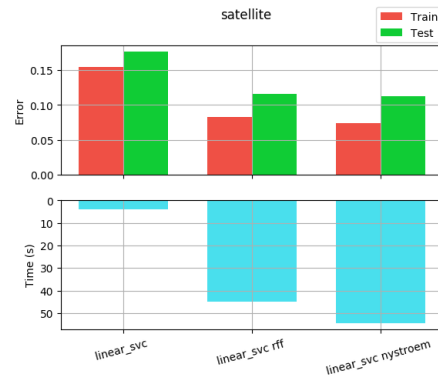


FIGURE F.6: Linear-SVC with RFF and Nyström

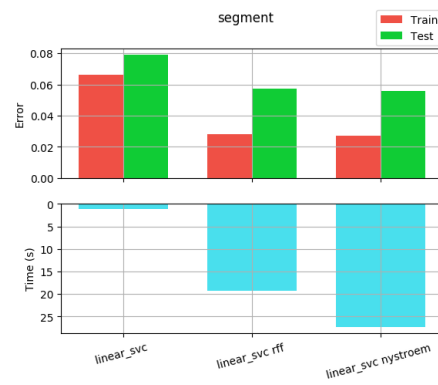


FIGURE F.7: Linear-SVC with RFF and Nyström

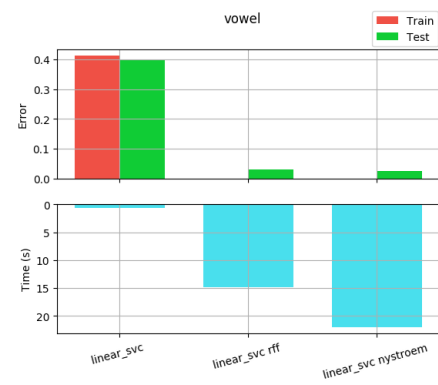


FIGURE F.8: Linear-SVC with RFF and Nyström

Appendix G

Results of experiment 2.6

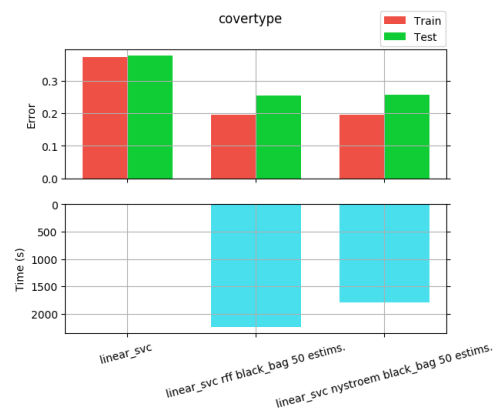


FIGURE G.1: Linear-SVM with Black Bag model

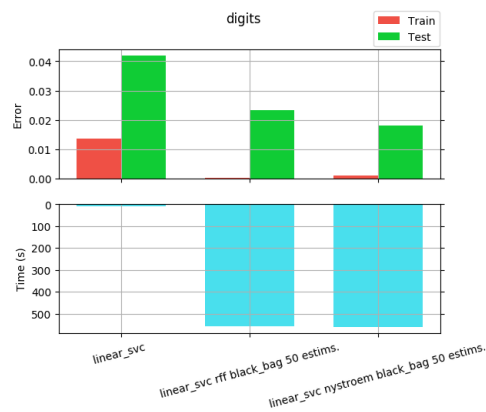


FIGURE G.2: Linear-SVM with Black Bag model

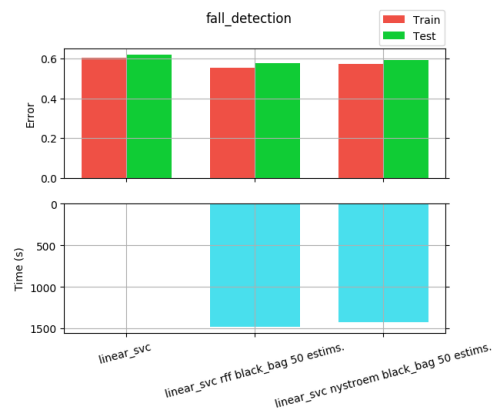


FIGURE G.3: Linear-SVM with Black Bag model

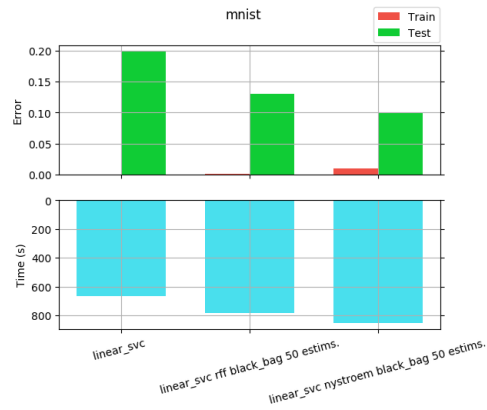


FIGURE G.4: Linear-SVM with Black Bag model

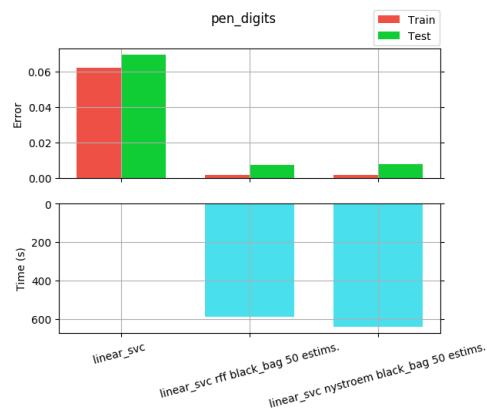


FIGURE G.5: Linear-SVM with Black Bag model

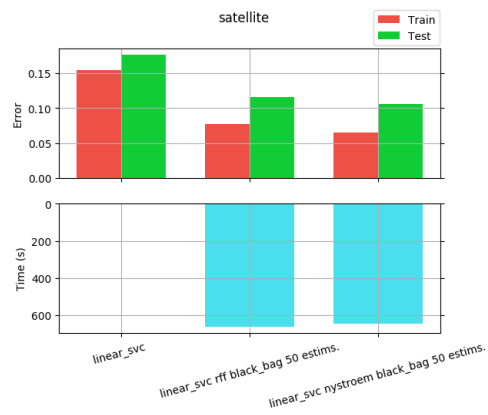


FIGURE G.6: Linear-SVM with Black Bag model

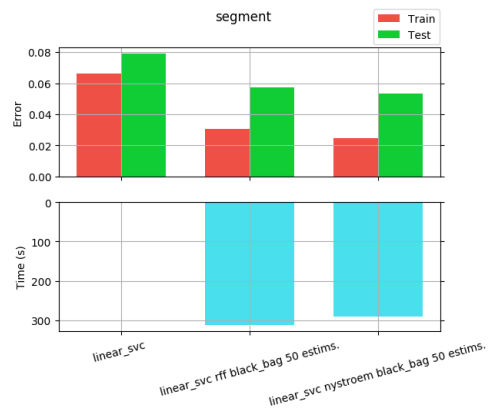


FIGURE G.7: Linear-SVM with Black Bag model

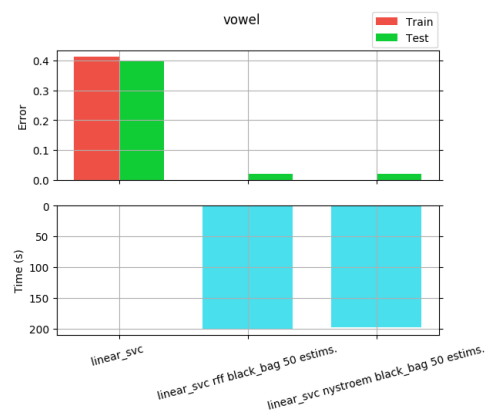


FIGURE G.8: Linear-SVM with Black Bag model

Appendix H

Results of experiment 2.2

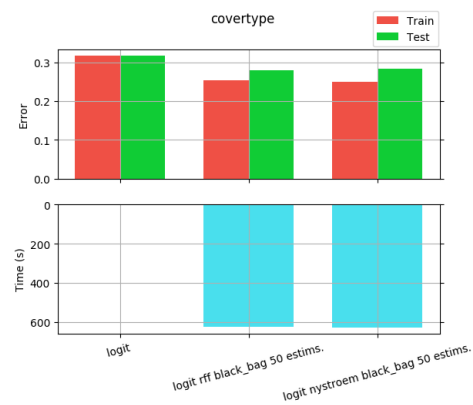


FIGURE H.1: Logistic Regression with Black Bag model

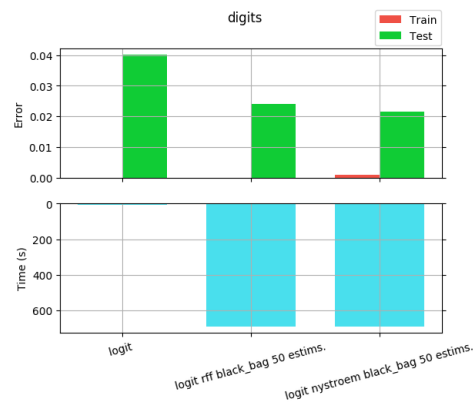


FIGURE H.2: Logistic Regression with Black Bag model

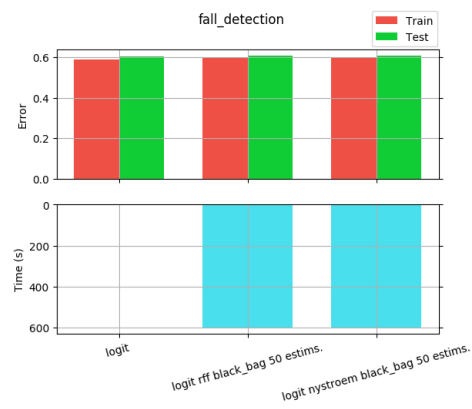


FIGURE H.3: Logistic Regression with Black Bag model

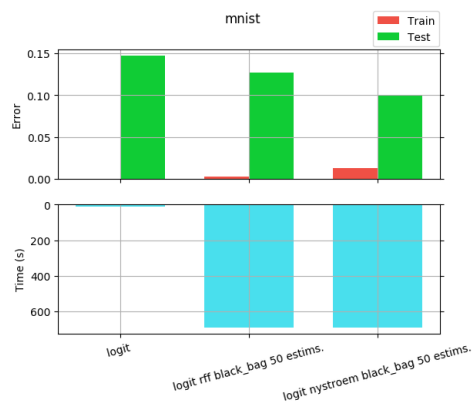


FIGURE H.4: Logistic Regression with Black Bag model

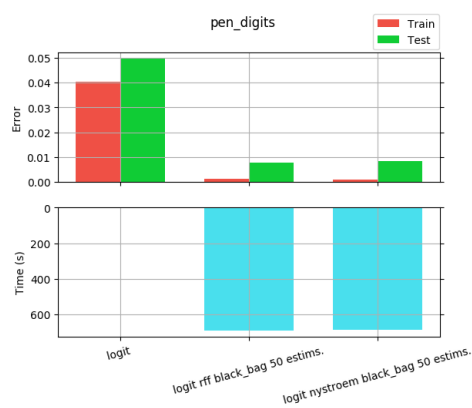


FIGURE H.5: Logistic Regression with Black Bag model

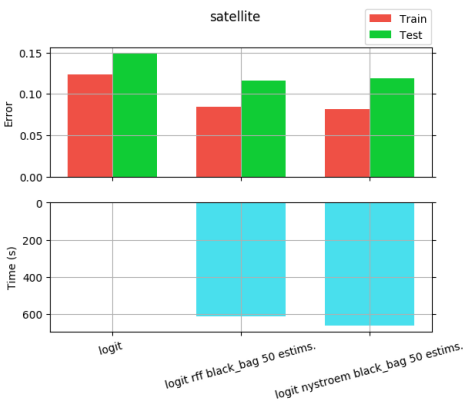


FIGURE H.6: Logistic Regression with Black Bag model

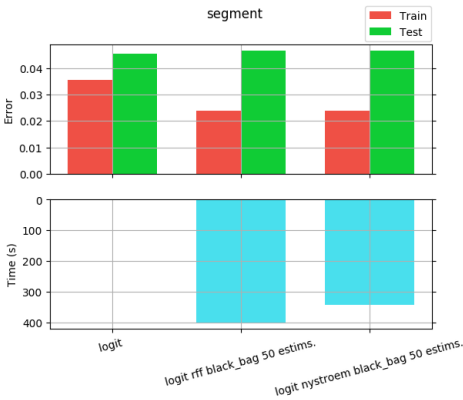


FIGURE H.7: Logistic Regression with Black Bag model

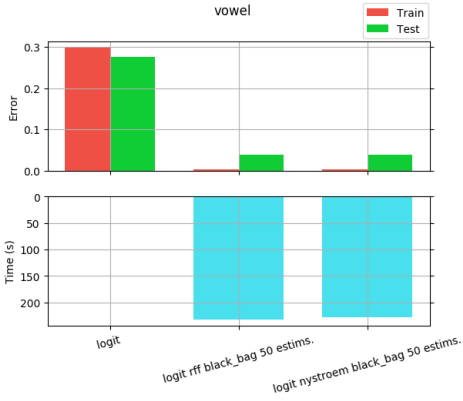


FIGURE H.8: Logistic Regression with Black Bag model

Appendix I

Results of experiment 2.2

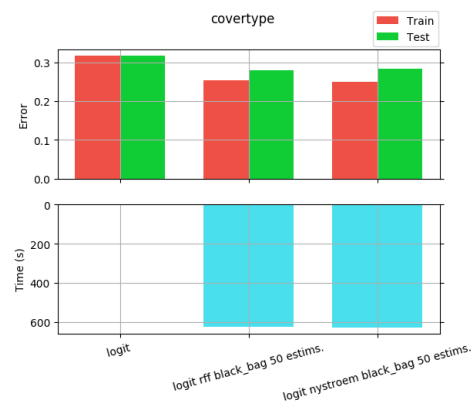


FIGURE I.1: Logistic Regression with Black Bag model

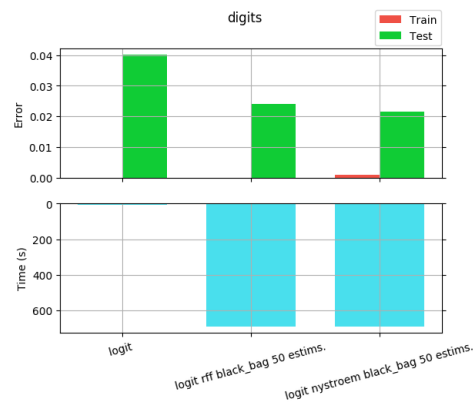


FIGURE I.2: Logistic Regression with Black Bag model

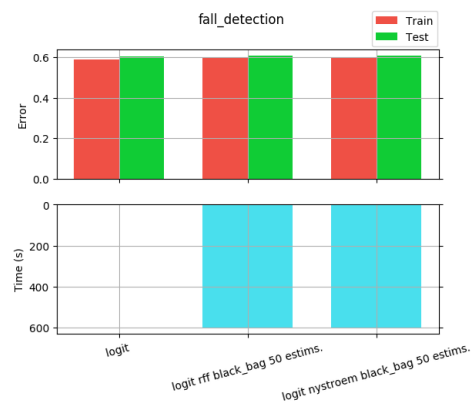


FIGURE I.3: Logistic Regression with Black Bag model

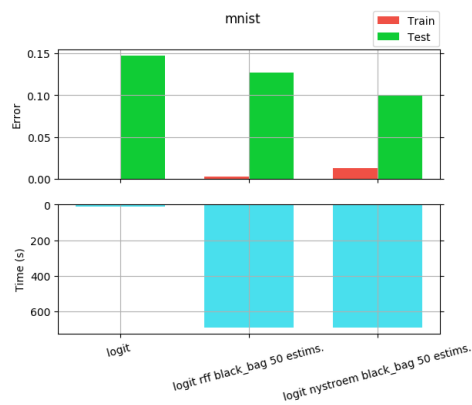


FIGURE I.4: Logistic Regression with Black Bag model

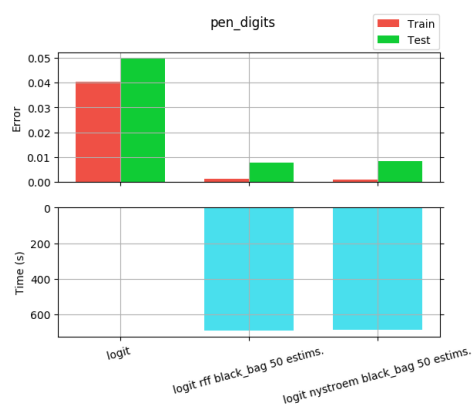


FIGURE I.5: Logistic Regression with Black Bag model

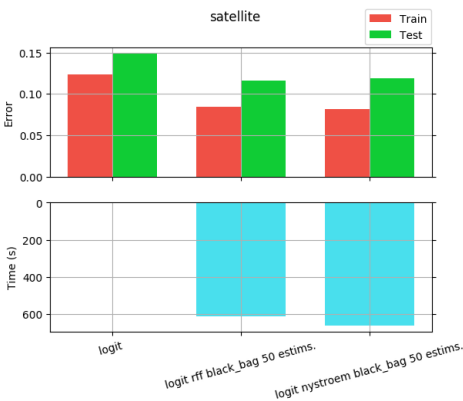


FIGURE I.6: Logistic Regression with Black Bag model

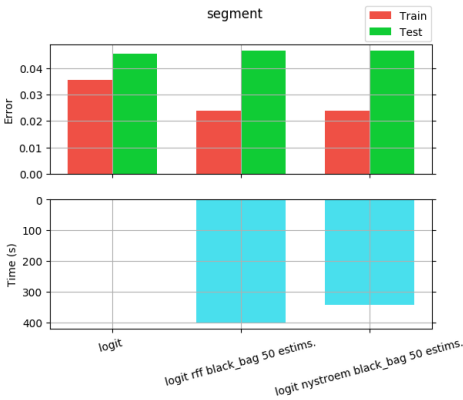


FIGURE I.7: Logistic Regression with Black Bag model

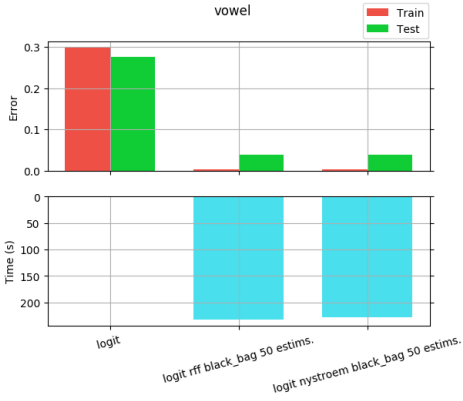


FIGURE I.8: Logistic Regression with Black Bag model