Handwritten number’s recognition

(Modellering av handskrivna siffror med maskininlärning)



Svetlana Oshcepkova

EC Utbildning

Machin learning project

March 2025

# Abstract

This project explores the classification of handwritten digits from the MNIST dataset using two machine learning models: Logistic Regression and Random Forest. The dataset is split into training and testing subsets. Hyperparameters are optimized via randomized search with cross-validation. The models are compared in terms of performance, and the best one is used for final predictions. Evaluation is conducted using a confusion matrix and single image prediction.

**Skapas automatiskt i Word genom att gå till Referenser > Innehållsförteckning.**

Innehållsförteckning

[Abstract 2](#_Toc160654338)

[Förkortningar och Begrepp Detta avsnitt behövs oftast inte. 3](#_Toc160654339)

[1 Inledning 1](#_Toc160654340)

[1.1 Underrubrik – Exempel 1](#_Toc160654341)

[2 Teori 2](#_Toc160654342)

[2.1 Exempel: Regressionsmodeller 2](#_Toc160654343)

[2.1.1 Exempel: Lasso 2](#_Toc160654344)

[2.1.2 Exempel: Ridge 2](#_Toc160654345)

[2.1.3 Exempel: Elastic Net 2](#_Toc160654346)

[2.2 Exempel: Neurala Nätverk 2](#_Toc160654347)

[3 Metod 3](#_Toc160654348)

[4 Resultat och Diskussion 4](#_Toc160654349)

[5 Slutsatser 5](#_Toc160654350)

[6 Teoretiska frågor 6](#_Toc160654351)

[7 Självutvärdering 7](#_Toc160654352)

[Appendix A 8](#_Toc160654353)

[Källförteckning 9](#_Toc160654354)

# Inledning

Syftet med denna rapport är att undersöka hur väl två klassiska modeller – Logistic Regression och Random Forest – kan klassificera handskrivna siffror från MNIST-datat.

För att uppfylla syftet kommer följande frågeställningar att besvaras:

1. Vilken av modellerna presterar bäst vid användning av cross-validation?
2. Hur bra förutsäger den bästa modellen siffror i testdatan?
3. Hur tolkar man resultatet i en confusion matrix?

## Underrubrik – Exempel

Text är skriven på formatet Calibri med textstorlek 11.

# Teori

2.1 Logistic Regression

### En klassisk linjär modell som ofta används för binär klassificering men kan anpassas för flerklasser (t.ex. one-vs-rest).

### Random Forest En ensemblemodell som bygger flera beslutsträd och kombinerar deras resultat för att förbättra prediktionens noggrannhet och minska överanpassning.

### Exempel: Elastic Net

## Exempel: Hyperparameteroptimering

## RandomizedSearchCV används för att hitta optimala kombinationer av hyperparametrar över ett fördefinierat område.

## Cross validation

K-fold cross-validation används för att robust utvärdera modeller under träning, vilket minskar risken för överanpassning.

3.Metod

* **Data:** MNIST (nMist\_784), innehåller 784 features (28x28 pixlar) och en målsiffra (0–9).
* **Datadelning:** Datasetet delas i träning (80%) och test (20%).
* **Modeller:** LogisticRegression och RandomForestClassifier från scikit-learn.
* **Optimering:** RandomizedSearchCV med cross\_val\_score används på träningsdata.
* **Utvärdering:** Bästa modellen testas på testdatan. Prestanda analyseras med accuracy score och confusion matrix.
* **Visualisering:** En slumpmässigt vald bild från testdatan används för att demonstrera modellens prediktion visuellt.

1. Resultat

|  |  |  |
| --- | --- | --- |
| Modell | Cross-val Accuracy | Test Accuracy |
| Logistic Regression |  |  |
| Random Forest classifier |  |  |

Tabell 1:Resultat av hyperparameteroptimerade modeller.

**Confusion matrix:** Random Forest visar tydligare klassificeringsnoggrannhet, med färre felklassificeringar bland liknande siffror (t.ex. 3 och 5).

**Visualisering:** En slumpmässig bild av siffran 8 klassificerades korrekt av Random Forest.

# Metod

Hur har du genomfört ditt arbete? Exempelvis, hur har datan erhållits?

# Resultat och Diskussion

|  |  |
| --- | --- |
| **RMSE för olika modeller** | |
| Enkel Linjär Regression | xx |
| Lasso | xx |
| Ridge | xx |

Tabell 1: Root Mean Squared Error (RMSE) för de fyra valda modellerna.

En bild som visar text, Teckensnitt, skärmbild, linje

Automatiskt genererad beskrivning

Figur 1: Hur man lägger in tabell eller figur nummer samt beskrivning.

# Slutsatser

Här besvarar du bl.a. frågeställningarna.

# Teoretiska frågor

# Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.
2. Vilket betyg du anser att du skall ha och varför.
3. Något du vill lyfta fram till Antonio?

# Appendix A

# Källförteckning