Supervised Learning, Part 2: Classification

Elliott Ash

Max Planck Summer School 2017

Regression vs. Classification

- In economics and other social sciences, most empirical analysis has an ordered, real-valued or binary (one-dimensional), outcome.
  - Regression gives us a linear prediction of $Y \in \mathbb{R}$ given the predictors X.
- But what if the outcome Y is a multi-dimensional classification, rather than a number?
  - For example, deciding the legal area of a court case, when there are dozens of areas.
    - There's no way to line legal areas up in one dimension.
  - This is a **classification task** rather than a regression task.

Regression vs. Classification

- In economics and other social sciences, most empirical analysis has an ordered, real-valued or binary (one-dimensional), outcome.
  - Regression gives us a linear prediction of $Y \in \mathbb{R}$ given the predictors X.
- But what if the outcome Y is a multi-dimensional classification, rather than a number?
  - For example, deciding the legal area of a court case, when there are dozens of areas.
    - There's no way to line legal areas up in one dimension.
  - This is a **classification task** rather than a regression task.

Regression vs. Classification

- In economics and other social sciences, most empirical analysis has an ordered, real-valued or binary (one-dimensional), outcome.
  - Regression gives us a linear prediction of $Y \in \mathbb{R}$ given the predictors X.
- But what if the outcome Y is a multi-dimensional classification, rather than a number?
  - For example, deciding the legal area of a court case, when there are dozens of areas.
    - There's no way to line legal areas up in one dimension.
  - This is a **classification task** rather than a regression task.

Regression vs. Classification

- In economics and other social sciences, most empirical analysis has an ordered, real-valued or binary (one-dimensional), outcome.
  - Regression gives us a linear prediction of $Y \in \mathbb{R}$ given the predictors X.
- But what if the outcome Y is a multi-dimensional classification, rather than a number?
  - For example, deciding the legal area of a court case, when there are dozens of areas.
    - There's no way to line legal areas up in one dimension.
  - This is a **classification task** rather than a regression task.

Regression vs. Classification

- In economics and other social sciences, most empirical analysis has an ordered, real-valued or binary (one-dimensional), outcome.
  - Regression gives us a linear prediction of $Y \in \mathbb{R}$ given the predictors X.
- But what if the outcome Y is a multi-dimensional classification, rather than a number?
  - For example, deciding the legal area of a court case, when there are dozens of areas.
    - There's no way to line legal areas up in one dimension.
  - This is a **classification task** rather than a regression task.

- Say you have a set of documents and a set of legal areas.
  - You only know the legal area for a subset of documents which have been hand-coded.
  - Can we use the existing labels, and the text of all cases, to machine-code the labels for the unlabelled cases?

Document Classification

- Say you have a set of documents and a set of legal areas.
  - You only know the legal area for a subset of documents which have been hand-coded.
  - Can we use the existing labels, and the text of all cases, to machine-code the labels for the unlabelled cases?

Document Classification

- Say you have a set of documents and a set of legal areas.
  - You only know the legal area for a subset of documents which have been hand-coded.
  - Can we use the existing labels, and the text of all cases, to machine-code the labels for the unlabelled cases?

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks
2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks
2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Regularized Multinomial Logistic Regression

- Regularized logistic regression is a popular choice for classification to multiple discrete categories.
- The model estimates the parameters β for

$$\Pr(Y_i = c) = \frac{e^{\beta_c X_i}}{\sum_{k=1}^{K} e^{\beta_k X_i}}$$

  for a set of class labels $c \in \{1, ..., K\}$.
- Coefficients are estimated with gradient-based stochastic optimization routines, which approximate regularized MLE.
- [demo_code.py]

Regularized Multinomial Logistic Regression

- Regularized logistic regression is a popular choice for classification to multiple discrete categories.
- The model estimates the parameters β for

$$\Pr(Y_i = c) = \frac{e^{\beta_c X_i}}{\sum_{k=1}^{K} e^{\beta_k X_i}}$$

  for a set of class labels $c \in \{1, ..., K\}$.
- Coefficients are estimated with gradient-based stochastic optimization routines, which approximate regularized MLE.
- [demo_code.py]

Regularized Multinomial Logistic Regression

- Regularized logistic regression is a popular choice for classification to multiple discrete categories.
- The model estimates the parameters β for

$$\Pr(Y_i = c) = \frac{e^{\beta_c X_i}}{\sum_{k=1}^{K} e^{\beta_k X_i}}$$

  for a set of class labels $c \in \{1, ..., K\}$.

- Coefficients are estimated with gradient-based stochastic optimization routines, which approximate regularized MLE.
- [demo_code.py]

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks
2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Random Forests

- Random Forest Classifier provides an ensemble of decision trees (therefore, forest):
  - The trees "vote" on the classification and the forest aggregates the votes.
- Good prediction performance – due to out-of-sample validation being included in the training process.
- Also, interpretable because includes a feature importance ranking.
- [demo_code.py]

Random Forests

- Random Forest Classifier provides an ensemble of decision trees (therefore, forest):
  - The trees "vote" on the classification and the forest aggregates the votes.
- Good prediction performance – due to out-of-sample validation being included in the training process.
- Also, interpretable because includes a feature importance ranking.
- [demo_code.py]

Random Forests

- Random Forest Classifier provides an ensemble of decision trees (therefore, forest):
  - The trees "vote" on the classification and the forest aggregates the votes.
- Good prediction performance – due to out-of-sample validation being included in the training process.
- Also, interpretable because includes a feature importance ranking.
- [demo_code.py]

Random Forests

- Random Forest Classifier provides an ensemble of decision trees (therefore, forest):
  - The trees “vote” on the classification and the forest aggregates the votes.
- Good prediction performance – due to out-of-sample validation being included in the training process.
- Also, interpretable because includes a feature importance ranking.
- [demo_code.py]

Boosted Trees

- XGBoost is another ensemble method with proven performance. Details and code are online.

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks

2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Deep Neural Networks (DNNs)

- DNNs are the new frontier for prediction with high-dimensional data.
- They do a better job than other models in recovering interaction effects between predictors.
- See the Python **keras** package.

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks

2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks
2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Classifying Policy Topics in Political Texts

- This paper develops a new method for supervised learning of policy topics, as well as support for those topics, in political texts.
- Data:
  - Download the hand-coded party platforms from the Manifestos project
    - Only English-language platforms (for now)
  - Corpus has 52,151 annotated statements.

Manifestos Data (36 policy topics)

Code	Policy Topic	Valence
101	foreign special	+
102	foreign special	-
103	anti-imperalism	
104	military	+
105	military	-
106	peace	
107	internationalism	+
108	europe	+
109	internationalism	-
110	europe	-
201	freedom & human rights	
202	democracy	
203	constitution	+
204	constitution	-
...	...	...

Model Training

1. Train a classifier to learn the policy topic from the text of a statement.
   - We got best results with logistic regression.
   - About 50% of the topics are correctly coded, which is very high given there are 36 topics.
2. Train a regression model to learn the valence from the text of a statement
   - we used a random forest regressor to learn the valence, with correlation between truth and prediction of .8

Model Training

1. Train a classifier to learn the policy topic from the text of a statement.
   - We got best results with logistic regression.
   - About 50% of the topics are correctly coded, which is very high given there are 36 topics.
2. Train a regression model to learn the valence from the text of a statement
   - we used a random forest regressor to learn the valence, with correlation between truth and prediction of .8

Model Training

1. Train a classifier to learn the policy topic from the text of a statement.
   - We got best results with logistic regression.
   - About 50% of the topics are correctly coded, which is very high given there are 36 topics.
2. Train a regression model to learn the valence from the text of a statement
   - we used a random forest regressor to learn the valence, with correlation between truth and prediction of .8

Model Training

1. Train a classifier to learn the policy topic from the text of a statement.
    - We got best results with logistic regression.
    - About 50% of the topics are correctly coded, which is very high given there are 36 topics.
2. Train a regression model to learn the valence from the text of a statement
    - we used a random forest regressor to learn the valence, with correlation between truth and prediction of .8

Model Training

1. Train a classifier to learn the policy topic from the text of a statement.
    - We got best results with logistic regression.
    - About 50% of the topics are correctly coded, which is very high given there are 36 topics.
2. Train a regression model to learn the valence from the text of a statement
    - we used a random forest regressor to learn the valence, with correlation between truth and prediction of .8

Application of model

- We use this model to examine the impact of a natural experiment in New Zealand:
  - 1993 reform from first-past-the-post (majoritarian) elections to mixed-member proportional representation.
  - We look at changes in ideological positioning between parties due to the reform using the text of their parliamentary speeches.
  - the policy-topics model is used to understand better on what policies they are changing positions (still preliminary).

Application of model

- We use this model to examine the impact of a natural experiment in New Zealand:
  - 1993 reform from first-past-the-post (majoritarian) elections to mixed-member proportional representation.
  - We look at changes in ideological positioning between parties due to the reform using the text of their parliamentary speeches.
  - the policy-topics model is used to understand better on what policies they are changing positions (still preliminary).

Application of model

- We use this model to examine the impact of a natural experiment in New Zealand:
  - 1993 reform from first-past-the-post (majoritarian) elections to mixed-member proportional representation.
  - We look at changes in ideological positioning between parties due to the reform using the text of their parliamentary speeches.
  - the policy-topics model is used to understand better on what policies they are changing positions (still preliminary).

Application of model

- We use this model to examine the impact of a natural experiment in New Zealand:
  - 1993 reform from first-past-the-post (majoritarian) elections to mixed-member proportional representation.
  - We look at changes in ideological positioning between parties due to the reform using the text of their parliamentary speeches.
  - the policy-topics model is used to understand better on what policies they are changing positions (still preliminary).

Outline

1. Classification Models
   - Regularized Multinomial Logistic Regression
   - Ensemble Methods: Random Forests and XGBoost
   - Deep Neural Networks
2. Applications
   - Ash, Morelli, and Osnabruegge (2017)
   - Jelveh, Kogut, and Naidu (2016)

Political language in economics

- This paper looks at political language and ideology in the economics literature.
- They use data on campaign contributions to assign a subset of economists to Republican or Democrat.
- Then they train a classifier to predict party based on the text of written articles
  - They use an ensemble PLS model, that "votes" in the same way as random forests, but the constituent voters are PLS regressors, rather than decision trees.
  - They control for topic choices using JEL K codes and LDA topics.
  - The model predicts with 70% accuracy.

Political language in economics

- This paper looks at political language and ideology in the economics literature.
- They use data on campaign contributions to assign a subset of economists to Republican or Democrat.
- Then they train a classifier to predict party based on the text of written articles
  - They use an ensemble PLS model, that "votes" in the same way as random forests, but the constituent voters are PLS regressors, rather than decision trees.
  - They control for topic choices using JEL K codes and LDA topics.
  - The model predicts with 70% accuracy.

Political language in economics

- This paper looks at political language and ideology in the economics literature.
- They use data on campaign contributions to assign a subset of economists to Republican or Democrat.
- Then they train a classifier to predict party based on the text of written articles
  - They use an ensemble PLS model, that "votes" in the same way as random forests, but the constituent voters are PLS regressors, rather than decision trees.
  - They control for topic choices using JEL K codes and LDA topics.
  - The model predicts with 70% accuracy.

Political language in economics

- This paper looks at political language and ideology in the economics literature.
- They use data on campaign contributions to assign a subset of economists to Republican or Democrat.
- Then they train a classifier to predict party based on the text of written articles
  - They use an ensemble PLS model, that "votes" in the same way as random forests, but the constituent voters are PLS regressors, rather than decision trees.
  - They control for topic choices using JEL K codes and LDA topics.
  - The model predicts with 70% accuracy.

Political language in economics

- This paper looks at political language and ideology in the economics literature.
- They use data on campaign contributions to assign a subset of economists to Republican or Democrat.
- Then they train a classifier to predict party based on the text of written articles
  - They use an ensemble PLS model, that "votes" in the same way as random forests, but the constituent voters are PLS regressors, rather than decision trees.
  - They control for topic choices using JEL K codes and LDA topics.
  - The model predicts with 70% accuracy.

Political language in economics

- This paper looks at political language and ideology in the economics literature.
- They use data on campaign contributions to assign a subset of economists to Republican or Democrat.
- Then they train a classifier to predict party based on the text of written articles
  - They use an ensemble PLS model, that "votes" in the same way as random forests, but the constituent voters are PLS regressors, rather than decision trees.
  - They control for topic choices using JEL K codes and LDA topics.
  - The model predicts with 70% accuracy.

Political Language in Economics: Results

- Jelveh et al (2016) then demonstrate significant ideological sorting across fields:
  - law and economics is the most-right wing field, labor economics is the most left-wing field
- Right-wing economists report a higher labor supply elasticity than left-wing economists
- The ideology of editors does not affect ideology of published articles.

Political Language in Economics: Results

- Jelveh et al (2016) then demonstrate significant ideological sorting across fields:
  - law and economics is the most-right wing field, labor economics is the most left-wing field
- Right-wing economists report a higher labor supply elasticity than left-wing economists
- The ideology of editors does not affect ideology of published articles.

Political Language in Economics: Results

- Jelveh et al (2016) then demonstrate significant ideological sorting across fields:
  - law and economics is the most-right wing field, labor economics is the most left-wing field
- Right-wing economists report a higher labor supply elasticity than left-wing economists
- The ideology of editors does not affect ideology of published articles.