Machine Learning for Official Statistics and SDGs

Supervised Classification

UNITED NATIONS
SIAP
Statistical Institute for Asia and the Pacific

[Classification]

What is a classification problem?

[Classification]

What is a classification problem?

- The goal of classification is to understand why an observation belongs to a certain category

[Classification]

What is a classification problem?

- The goal of classification is to understand why an observation belongs to a certain category
- The interest is on y that takes discrete values: 0/1, high school/prilmary school/no education; urban/rural

[Classification]

What is a classification problem?

- The goal of classification is to understand why an observation belongs to a certain category
- The interest is on y that takes discrete values: 0/1, high school/prilmary school/no education; urban/rural
- There may be variables explaining why we observe that y belongs to a particular category

[Classification]

What is a classification problem?

- The goal of classification is to understand why an observation belongs to a certain category
- The interest is on y that takes discrete values: 0/1, high school/prilmary school/no education; urban/rural
- There may be variables explaining why we observe that y belongs to a particular category
- A classifier is a tool that provides a classification for y

[Supervised *vs* unsupervised Classification]

[SUPERVISED *vs* UNSUPERVISED CLASSIFICATION]

- In **supervised** classification, we observe the class for y

[Supervised *vs* unsupervised Classification]

- In **supervised** classification, we observe the class for y

  One may learn from that information and estimate the impact of other variables on that classification (*e.g.* logit regression)

[Supervised *vs* unsupervised Classification]

- In **supervised** classification, we observe the class for y

  One may learn from that information and estimate the impact of other variables on that classification (*e.g*. logit regression)
- In **unsupervised** classification, we observe a set of variables for each observation

[Supervised *vs* unsupervised Classification]

- In **supervised** classification, we observe the class for y

  One may learn from that information and estimate the impact of other variables on that classification (*e.g*. logit regression)

- In **unsupervised** classification, we observe a set of variables for each observation

  The goal is to classify observations from those variables (clustering) without having any information of what a category means.

[Supervised *vs* unsupervised Classification]

- In **supervised** classification, we observe the class for y

  One may learn from that information and estimate the impact of other variables on that classification (*e.g.* logit regression)
- In **unsupervised** classification, we observe a set of variables for each observation

  The goal is to classify observations from those variables (clustering) without having any information of what a category means.

[SUPERVISED *vs* UNSUPERVISED CLASSIFICATION]

- In **supervised** classification, we observe the class for y

  One may learn from that information and estimate the impact of other variables on that classification (*e.g*. logit regression)
- In **unsupervised** classification, we observe a set of variables for each observation

  The goal is to classify observations from those variables (clustering) without having any information of what a category means.

- We'll focus on **supervised** classification

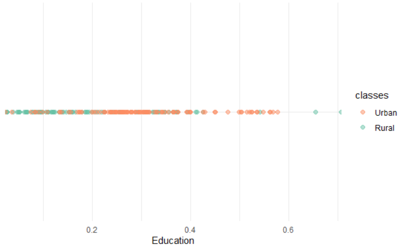
[Classification: An example]

[Classification: An example]

- You observe households that are either in *Urban* or *Rural* areas (colors) and one variable (feature): *Education*.

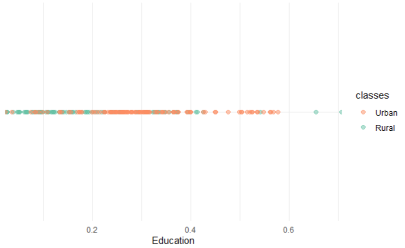
[Classification: An example]

- You observe households that are either in *Urban* or *Rural* areas (colors) and one variable (feature): *Education*.



[Classification: An example]

- You observe households that are either in *Urban* or *Rural* areas (colors) and one variable (feature): *Education*.



- A classifier determines the value of *Education* that separate "Rural" from "Urban", typically with a threshold rule : "if $x \geq x_t$ then y is categorized as Urban"

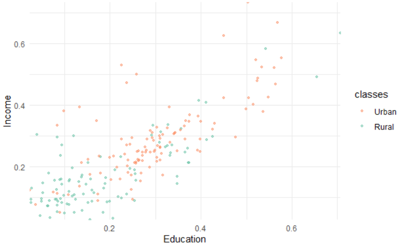
[Classification: A 2-D example]

[Classification: A 2-D example]

- You observe households that are either in *Urban* or *Rural* areas (colors) and **two** variables (features): *Education* and *Income*

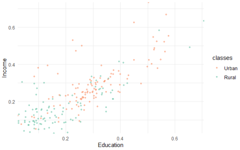
[Classification: A 2-D example]

- You observe households that are either in *Urban* or *Rural* areas (colors) and **two** variables (features): *Education* and *Income*

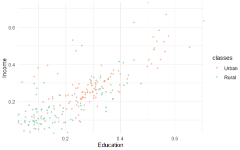


[Classification: A 2-D example]

[Classification: A 2-D example]

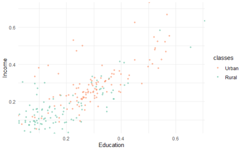


[Classification: A 2-D example]



- A classifier will determine a **boundary** using both *Education* and *Income* to separate "Rural" from "Urban"

[CLASSIFICATION: A 2-D EXAMPLE]



- A classifier will determine a **boundary** using both *Education* and *Income* to separate "Rural" from "Urban"
- The rule can be based on a linear relationship between *Education* and *Income* or can be non linear.

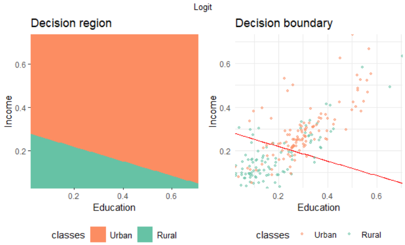
[Classification: A 2-D example]

[Classification: A 2-D example]

- The logit is an example of linear classifier

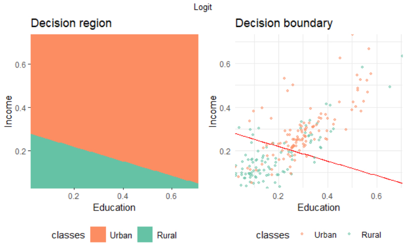
[Classification: A 2-D example]

- The logit is an example of linear classifier



[CLASSIFICATION: A 2-D EXAMPLE]

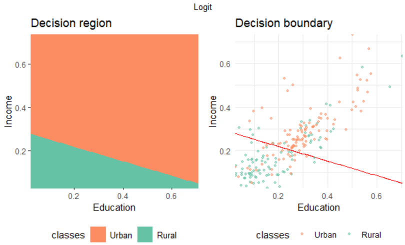
- The logit is an example of linear classifier



- The rule that separated the two classes is $x'\beta \geq T_0$ with T_0 a known threshold

[Classification: A 2-D example]

- The logit is an example of linear classifier



- The rule that separated the two classes is $x'\beta \geq T_0$ with T_0 a known threshold

  e.g. $\beta_1 Education + \beta_2 Income \geq T_0$

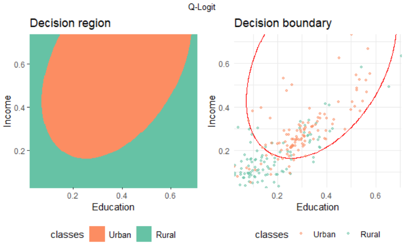
[Classification: A 2-D example]

[Classification: A 2-D example]

- The quadratic logit is an example of non-linear classifier

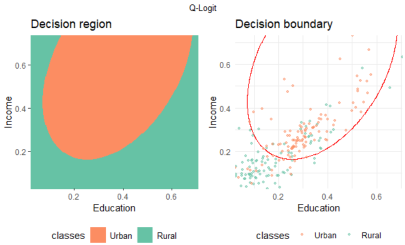
[Classification: A 2-D example]

- The quadratic logit is an example of non-linear classifier



[Classification: A 2-D example]

- The quadratic logit is an example of non-linear classifier



- The rule that separated the two classes is non linear in the variables *Education* and *Income*

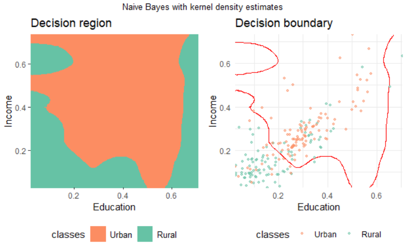
[Classification: A 2-D example]

[Classification: A 2-D example]

- Other examples can be very non linear

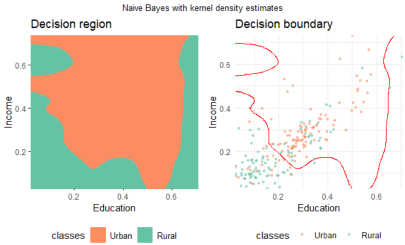
[Classification: A 2-D example]

- Other examples can be very non linear



[CLASSIFICATION: A 2-D EXAMPLE]

- Other examples can be very non linear



- It is hard to understand how the two classes are built using *Education* and *Income*

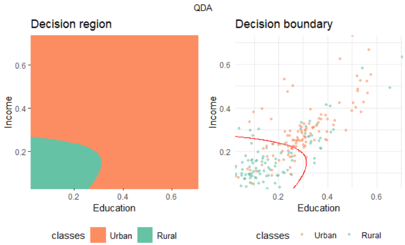
[Classification: A 2-D example]

[Classification: A 2-D example]

- Other examples can be very non linear

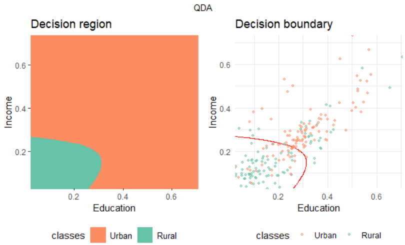
[Classification: A 2-D example]

- Other examples can be very non linear



[Classification: A 2-D example]

- Other examples can be very non linear



- The boundary is very complex but uses *Education* and *Income* features.

[How to select the right model?]

[How to select the right model?]

- What is the goal?

[How to select the right model?]

- What is the goal?
  Have the “best‘” classification

[How to select the right model?]

- What is the goal?

  Have the “best‘” classification

$\hookrightarrow$ Need for a criterion to determine what is a good classifier

[How to select the right model?]

- What is the goal?
  Have the “best‘” classification

$\hookrightarrow$ Need for a criterion to determine what is a good classifier

- Measures of fit in classification are different and specific

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

- Accuracy

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

- Accuracy
- Confusion matrix

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

- Accuracy
- Confusion matrix
- Specificty & sensitivity

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

- Accuracy
- Confusion matrix
- Specificty & sensitivity
- Kappa

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

- Accuracy
- Confusion matrix
- Specificty & sensitivity
- Kappa

...

[Measures of fit in classification]

There are several popular measures of fit, differing in their spirit and their goal

- Accuracy
- Confusion matrix
- Specificty & sensitivity
- Kappa

  ...

  All answer to different questions and solve different problems

[Accuracy and Confusion Matrix]

Accuracy corresponds to the probability of being "accurate"

$$\Pr\left[y_0 = \widehat{f}(x_0)\right]$$

[Accuracy and Confusion Matrix]

Accuracy corresponds to the probability of being "accurate"

$$\Pr\left[y_0 = \widehat{f}(x_0)\right]$$

- where $\widehat{f}(\cdot)$ is the classifier.

[Accuracy and Confusion Matrix]

Accuracy corresponds to the probability of being "accurate"

$$\Pr\left[y_0 = \widehat{f}(x_0)\right]$$

- where $\widehat{f}(\cdot)$ is the classifier.

$\hookrightarrow$ We want the maximum possible accuracy.

[Accuracy and Confusion Matrix]

Accuracy corresponds to the probability of being "accurate"

$$\Pr\left[y_0 = \widehat{f}(x_0)\right]$$

- where $\widehat{f}(\cdot)$ is the classifier.

$\hookrightarrow$ We want the maximum possible accuracy.

- Equivalently, we may want to minimize the *error rate* or *misclassification rate*

$$\Pr\left[y_0 \neq \widehat{f}(x_0)\right]$$

[Confusion Matrix & Accuracy]

A classifier predicts in which class each observation should be:

[Confusion Matrix & Accuracy]

A classifier predicts in which class each observation should be:

	Observed (True)	
Predicted	TP (True Positive)	FP (False Positive)
	FN (False Positive)	TN (True Negative)

Table: Confusion Matrix

[Confusion Matrix & Accuracy]

A classifier predicts in which class each observation should be:

	Observed (True)	
Predicted	TP (True Positive)	FP (False Positive)
	FN (False Positive)	TN (True Negative)

Table: Confusion Matrix

- Accuracy is then the ratio:

$$Acuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{TruePositives + TrueNegatives}{N}$$

[Confusion Matrix & Accuracy]

A classifier predicts in which class each observation should be:

	Observed (True)	
Predicted	TP (True Positive)	FP (False Positive)
	FN (False Positive)	TN (True Negative)

Table: Confusion Matrix

- Accuracy is then the ratio:

$$Acuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{TruePositives + TrueNegatives}{N}$$

- It is the proportion of accurate predictions

[Confusion Matrix & Accuracy]

In practice, with a classifier we have:

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

[Confusion Matrix & Accuracy]

In practice, with a classifier we have:

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- Remember that here *Urban* is the "positive" class

[Confusion Matrix & Accuracy]

In practice, with a classifier we have:

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- Remember that here *Urban* is the "positive" class
- Accuracy is then the ratio:

$$Acuracy = \frac{87+69}{87+69+28+24}$$

$$= \frac{156}{208} = 0.75$$

[Confusion Matrix & Accuracy]

In practice, with a classifier we have:

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- Remember that here *Urban* is the "positive" class
- Accuracy is then the ratio:

$$Acuracy = \frac{87 + 69}{87 + 69 + 28 + 24}$$
$$= \frac{156}{208} = 0.75$$

- We have an accurate prediction in 75% of the cases.

[Confusion Matrix & Accuracy]

Accuracy is not the panacea and may be misleading

Pb 1: Imagine a data set where 95% of the observations are in class 1 and 5% in the remaining class

[Confusion Matrix & Accuracy]

Accuracy is not the panacea and may be misleading

Pb 1: Imagine a data set where 95% of the observations are in class 1 and 5% in the remaining class

$\hookrightarrow$ A classifier that predicts always class 1 will have an accuracy of 95%!

[Confusion Matrix & Accuracy]

Accuracy is not the panacea and may be misleading

Pb 1: Imagine a data set where 95% of the observations are in class 1 and 5% in the remaining class

$\hookrightarrow$ A classifier that predicts always class 1 will have an accuracy of 95%!

- Accuracy is not very useful with an imbalanced data set and one may prefer another indicator: *kappa*

[Confusion Matrix & Accuracy]

Accuracy is not the panacea and may be misleading

Pb 1: Imagine a data set where 95% of the observations are in class 1 and 5% in the remaining class

$\hookrightarrow$ A classifier that predicts always class 1 will have an accuracy of 95%!

- Accuracy is not very useful with an imbalanced data set and one may prefer another indicator: *kappa*

Pb 2: One may be more interested in correctly predicting a particular outcome

[Confusion Matrix & Accuracy]

Accuracy is not the panacea and may be misleading

Pb 1: Imagine a data set where 95% of the observations are in class 1 and 5% in the remaining class

$\hookrightarrow$ A classifier that predicts always class 1 will have an accuracy of 95%!

- Accuracy is not very useful with an imbalanced data set and one may prefer another indicator: *kappa*

Pb 2: One may be more interested in correctly predicting a particular outcome

$\hookrightarrow$ This is often the case

[Confusion Matrix & Accuracy]

Accuracy is not the panacea and may be misleading

Pb 1: Imagine a data set where 95% of the observations are in class 1 and 5% in the remaining class

$\hookrightarrow$ A classifier that predicts always class 1 will have an accuracy of 95%!

- Accuracy is not very useful with an imbalanced data set and one may prefer another indicator: *kappa*

Pb 2: One may be more interested in correctly predicting a particular outcome

$\hookrightarrow$ This is often the case

- One may need other measures, such as *Specificity* or *Sensitivity*

[Sensitivity or *True Positive Rate*]

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

[SENSITIVITY OR *True Positive Rate*]

		Observed (True) Urban	Observed (True) Rural
Predicted	Urban	87 (TP)	28 (FP)
Predicted	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- *Sensitivity* focuses on "positives" (here *Urban*), *i.e.* on predicted positives *vs* the observed positives

$$Sensitivity = \frac{TP}{TP + FN}$$

$$= \frac{87}{87 + 24} = 0.78$$

[Sensitivity or *True Positive Rate*]

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- *Sensitivity* focuses on "positives" (here *Urban*), *i.e.* on predicted positives *vs* the observed positives

$$Sensitivity = \frac{TP}{TP + FN}$$

$$= \frac{87}{87 + 24} = 0.78$$

- On *Urban*, we correctly predict in 78% of the cases

[Specificity or *True Negative Rate*]

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

[SPECIFICITY OR *True Negative Rate*]

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- *Sensitivity* focuses on negatives (*Rural*), *i.e.* on predicted positives *vs* the observed positives

$$Specificity = \frac{TN}{TN + FP}$$

$$= \frac{69}{69 + 28} = 0.71$$

[Specificity or *True Negative Rate*]

		Observed (True)	
		Urban	Rural
Predicted	Urban	87 (TP)	28 (FP)
	Rural	24 (FN)	69 (TN)

Table: Confusion Matrix

- *Sensitivity* focuses on negatives (*Rural*), *i.e.* on predicted positives *vs* the observed positives

$$Specificity = \frac{TN}{TN + FP}$$

$$= \frac{69}{69 + 28} = 0.71$$

- On *Rural*, we predict correctly in **only** 71% of the cases

[Confusion Matrix & Kappa (κ)]

Kappa (κ) is defined to measure the accuracy with imbalanced classes

Its formal definition is given by

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

[Confusion Matrix & Kappa (κ)]

Kappa (κ) is defined to measure the accuracy with imbalanced classes

Its formal definition is given by

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o is the current classifier accuracy which is compared here with the accuracy of an uniformed classifier P_e

[Confusion Matrix & Kappa (κ)]

Kappa (κ) is defined to measure the accuracy with imbalanced classes

Its formal definition is given by

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o is the current classifier accuracy which is compared here with the accuracy of an uniformed classifier P_e

P_e is the accuracy of a classifier that would operate purely by chance, using no information.

[Confusion Matrix & Kappa (κ)]

Kappa (κ) is defined to measure the accuracy with imbalanced classes

Its formal definition is given by

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o is the current classifier accuracy which is compared here with the accuracy of an uniformed classifier P_e

P_e is the accuracy of a classifier that would operate purely by chance, using no information.

- P_o is simple (here = 0.75) while P_e is more complex to compute.

[Confusion Matrix & Kappa (κ)]

Kappa (κ) is defined to measure the accuracy with imbalanced classes

Its formal definition is given by

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o is the current classifier accuracy which is compared here with the accuracy of an uniformed classifier P_e

P_e is the accuracy of a classifier that would operate purely by chance, using no information.

- P_o is simple (here = 0.75) while P_e is more complex to compute.
- The larger κ is, the better the model for a given distribution of classes in a data set

[Logit as You know it]

[Logit as You know it]

When dealing with a discrete outcome y we cannot use a *direct* linear relationship between y and the explanatory variables x (*i.e. Education, Income*)

[LOGIT AS YOU KNOW IT]

When dealing with a discrete outcome y we cannot use a *direct* linear relationship between y and the explanatory variables x (*i.e. Education*, *Income*)

- Logit aims at estimating the probabilities $\pi = Probablity[y = 1]$, which are continuous ($\in [0, 1]$)

[Logit as You know it]

When dealing with a discrete outcome y we cannot use a *direct* linear relationship between y and the explanatory variables x (*i.e. Education, Income*)

- Logit aims at estimating the probabilities $\pi = Probablity[y = 1]$, which are continuous ($\in [0, 1]$)
- The definition of the logit model is:

$$\pi = Pr\,(y = 1) = F(x'\beta) = \frac{1}{1 + \exp(-x'\beta)}$$

[Logit as You know it]

When dealing with a discrete outcome y we cannot use a *direct* linear relationship between y and the explanatory variables x (*i.e. Education, Income*)

- Logit aims at estimating the probabilities $\pi = Probablity[y = 1]$, which are continuous ($\in [0, 1]$)
- The definition of the logit model is:

$$\pi = Pr\,(y = 1) = F(x'\beta) = \frac{1}{1 + \exp(-x'\beta)}$$

- This can be transformed into:

$$\pi = \frac{exp(-x'\beta)}{1 + exp(-x'\beta)}$$

[Logit as You **Don't** know it]

[Logit as You Don't know it]

From this equation

$$\pi = \frac{exp(-x'\beta)}{1 + exp(-x'\beta)}$$

[Logit as You Don't know it]

From this equation

$$\pi = \frac{exp(-x'\beta)}{1 + exp(-x'\beta)}$$

one gets the **linear** nature of the logit:

$$log(\frac{\pi}{1-\pi}) = x'\beta$$

[LOGIT AS YOU **DON'T** KNOW IT]

From this equation

$$\pi = \frac{exp(-x'\beta)}{1 + exp(-x'\beta)}$$

one gets the **linear** nature of the logit:

$$log(\frac{\pi}{1-\pi}) = x'\beta$$

where $\frac{\pi}{1-\pi}$ is the odd ratio $\in [0, \infty]$ with values indicating high or low probability that $y = 1$

[LOGIT AS YOU **DON'T** KNOW IT]

From this equation

$$\pi = \frac{exp(-x'\beta)}{1 + exp(-x'\beta)}$$

one gets the **linear** nature of the logit:

$$log(\frac{\pi}{1-\pi}) = x'\beta$$

where $\frac{\pi}{1-\pi}$ is the odd ratio $\in [0, \infty]$ with values indicating high or low probability that $y = 1$

$\hookrightarrow$ *"The logit models log of odd ratios as linear in x"*

[Logit as a classifier]

[Logit as a classifier]

- Once estimated, $\widehat{\pi}_i$ provide a simple rule for classification

[Logit as a classifier]

- Once estimated, $\widehat{\pi}_i$ provide a simple rule for classification

$$\widehat{\pi}_i > t_0 \Leftrightarrow \widehat{y}_i = 1$$

[Logit as a classifier]

- Once estimated, $\widehat{\pi}_i$ provide a simple rule for classification

$$\widehat{\pi}_i > t_0 \Leftrightarrow \widehat{y}_i = 1$$

Where t_0 is a threshold provability, by default $1/2$.

[Logit as a classifier]

- Once estimated, $\widehat{\pi}_i$ provide a simple rule for classification

$$\widehat{\pi}_i > t_0 \Leftrightarrow \widehat{y}_i = 1$$

  Where t_0 is a threshold provability, by default $1/2$.
- If $t_0 = 1/2$ (default), then the rule is equivalent to:

$$x_i'\widehat{\beta} > 0 \Leftrightarrow \widehat{y}_i = 1$$

[Logit as a classifier]

- Once estimated, $\widehat{\pi}_i$ provide a simple rule for classification

$$\widehat{\pi}_i > t_0 \Leftrightarrow \widehat{y}_i = 1$$

  Where t_0 is a threshold provability, by default $1/2$.
- If $t_0 = 1/2$ (default), then the rule is equivalent to:

$$x_i'\widehat{\beta} > 0 \Leftrightarrow \widehat{y}_i = 1$$

- If $t_0 \neq 1/2$:

$$x_i'\widehat{\beta} > T_0 \Leftrightarrow \widehat{y}_i = 1$$

[Logit as a classifier]

- Once estimated, $\widehat{\pi}_i$ provide a simple rule for classification

$$\widehat{\pi}_i > t_0 \Leftrightarrow \widehat{y}_i = 1$$

Where t_0 is a threshold provability, by default 1/2.

- If $t_0 = 1/2$ (default), then the rule is equivalent to:

$$x_i'\widehat{\beta} > 0 \Leftrightarrow \widehat{y}_i = 1$$

- If $t_0 \neq 1/2$:

$$x_i'\widehat{\beta} > T_0 \Leftrightarrow \widehat{y}_i = 1$$

$\hookrightarrow$ The logit classifier depends on the linear combination of the x's

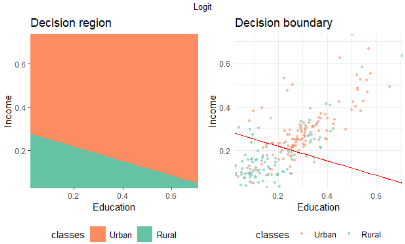
[Importance of the threshold]

[Importance of the threshold]

- The rule $x'\beta \geq T_0$ defines the partition of the space

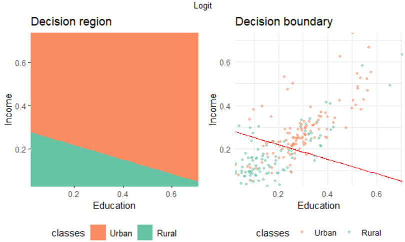
[Importance of the threshold]

- The rule $x'\beta \geq T_0$ defines the partition of the space



[Importance of the threshold]

- The rule $x'\beta \geq T_0$ defines the partition of the space



- This partition is sensitive to the choice of the threshold T_0 (and the t_0)

[Importance of the threshold]

[Importance of the threshold]

- Changing t_0 will change the predictions & the classification

[Importance of the threshold]

- Changing t_0 will change the predictions & the classification
  A higher t_0 will allocate less observations to the $y = 1$ category (Urban)

[Importance of the threshold]

- Changing t_0 will change the predictions & the classification

  A higher t_0 will allocate less observations to the $y = 1$ category (Urban)

  A lower t_0 will allocate more observations to the $y = 1$ category

[Importance of the threshold]

- Changing t_0 will change the predictions & the classification

  A higher t_0 will allocate less observations to the $y = 1$ category (Urban)

  A lower t_0 will allocate more observations to the $y = 1$ category
- The choice of t_o should be done according to the data and observed classes repartition

[IMPORTANCE OF THE THRESHOLD]

- Changing t_0 will change the predictions & the classification

  A higher t_0 will allocate less observations to the $y = 1$ category (Urban)

  A lower t_0 will allocate more observations to the $y = 1$ category
- The choice of t_o should be done according to the data and observed classes repartition
- *Specificity* and *Sensitivity* are affected by t_0

[The ROC Curve]

[The ROC Curve]

- We want the *Specificit*y and *Sensitivity* to be both **maximized** (ideally both would be 1)

[THE ROC CURVE]

- We want the *Specificit*y and *Sensitivity* to be both **maximized** (ideally both would be 1)
- The ROC curve help visualize the best choice

[The ROC Curve]

- We want the *Specificit*y and *Sensitivity* to be both **maximized** (ideally both would be 1)
- The ROC curve help visualize the best choice

  Be careful of the axes

[THE ROC CURVE]

- We want the *Specificit*y and *Sensitivity* to be both **maximized** (ideally both would be 1)
- The ROC curve help visualize the best choice
  Be careful of the axes
-

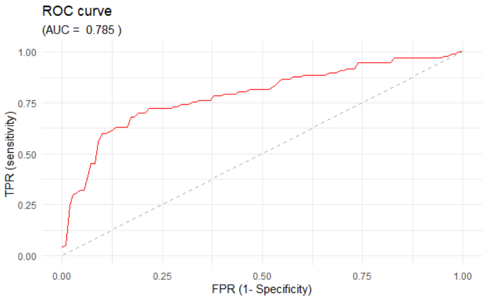
[THE ROC CURVE]

[The ROC Curve]

The ROC represents values of 1- Specificity = FPR *vs* Sensitivity = TPR for many values of the threshold t_0

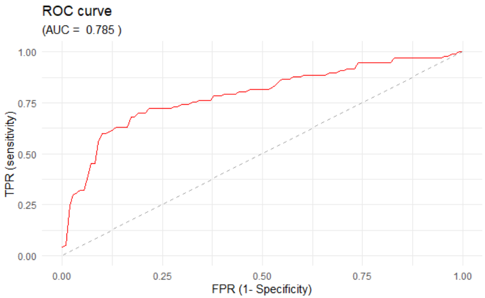
[The ROC Curve]

The ROC represents values of 1- Specificity = FPR *vs* Sensitivity = TPR for many values of the threshold t_0



[The ROC Curve]

The ROC represents values of 1- Specificity = FPR *vs* Sensitivity = TPR for many values of the threshold t_0



- *(sometimes x is sensitivity with inverted x-axis)*

[The ROC Curve: How to read?]

[The ROC Curve: How to read?]

Changing t_0 changes the classification of observations

[The ROC Curve: How to read?]

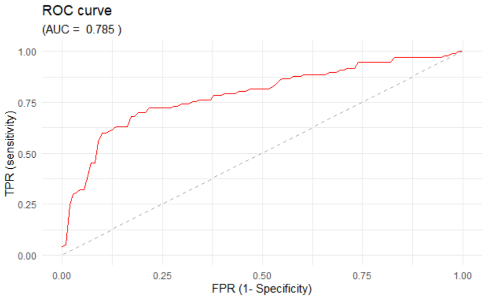
Changing t_0 changes the classification of observations

- Optimally, the curve should touch top-left corner

[The ROC Curve: How to read?]

Changing t_0 changes the classification of observations

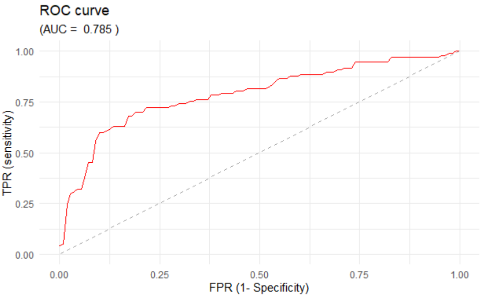
- Optimally, the curve should touch top-left corner



[The ROC Curve: How to read?]

Changing t_0 changes the classification of observations

- Optimally, the curve should touch top-left corner

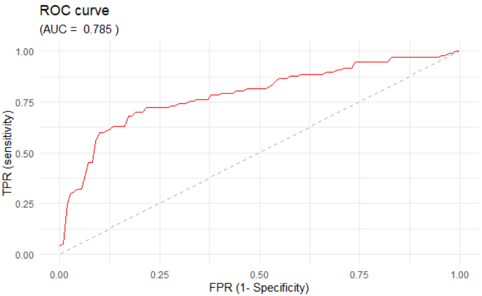


- If $t_0 \nearrow$, more cases classified as *Negatives*, less *Positives*

[The ROC Curve: How to read?]

Changing t_0 changes the classification of observations

- Optimally, the curve should touch top-left corner



- If $t_0 \nearrow$, more cases classified as *Negatives*, less *Positives*
- If $t_0 \nearrow$, specificity $\nearrow$ and sensitivity $\searrow$

[AUC as a measure of fit]

[Train & Validation set]

[CV on measures of fit]

[Model comparison]

[Takeways]

- In classification, the **Confusion matrix** is important

[Takeways]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.

[TAKEWAYS]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.
    - Sensitivity is accuracy restricted to the positives.

[TAKEWAYS]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.
  - Sensitivity is accuracy restricted to the positives.
  - Specificity is accuracy restricted to the negatives.

[Takeways]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.
  - Sensitivity is accuracy restricted to the positives.
  - Specificity is accuracy restricted to the negatives.
- When outcome is *imbalanced*, one may use **kappa** has a better measure for accuracy.

[TAKEWAYS]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.
  - Sensitivity is accuracy restricted to the positives.
  - Specificity is accuracy restricted to the negatives.
- When outcome is *imbalanced*, one may use **kappa** has a better measure for accuracy.

  Which measure you should consider depends on the context and your goal.

[Takeways]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.
  - Sensitivity is accuracy restricted to the positives.
  - Specificity is accuracy restricted to the negatives.
- When outcome is *imbalanced*, one may use **kappa** has a better measure for accuracy.

  Which measure you should consider depends on the context and your goal.
- **Logit** is a benchmark parametric model for classification

[Takeways]

- In classification, the **Confusion matrix** is important
- Many adjustment measures: **accuracy**, **sensitivity** and **specificity**.
  - Sensitivity is accuracy restricted to the positives.
  - Specificity is accuracy restricted to the negatives.
- When outcome is *imbalanced*, one may use **kappa** has a better measure for accuracy.

  Which measure you should consider depends on the context and your goal.
- **Logit** is a benchmark parametric model for classification

  One may use the **ROC** to change the threshold parameter

[Takeways]

- Use Training-Validation set to **select** parameters within a model

[Takeways]

- Use Training-Validation set to **select** parameters within a model
- Use Training-Validation set to **compare** models on the same criteria

[Takeways]

- Use Training-Validation set to **select** parameters within a model
- Use Training-Validation set to **compare** models on the same criteria
- Several criteria / measures of fit / cost functions are available

[TAKEWAYS]

- Use Training-Validation set to **select** parameters within a model
- Use Training-Validation set to **compare** models on the same criteria
- Several criteria / measures of fit / cost functions are available
- Time is the limit...