EVALUATING & INTERPRETING CLASSIFIERS

LECTURE 4

SECTION 3

JUNE 8TH

IACS
INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

UNIVERSITY of RWANDA

INTERPRETING LOGISTIC REGRESSION MODELS

INTERPRETING LOGISTIC REGRESSION MODEL:

The logistic regression model models the probability that a point x is labeled 1:

$$p(y=1|x,w)=\sigma(f_w(x))$$

The larger the value of $f_w(x)$ the higher the probability of $y=1$.

Linear Boundary

model for loan approval:

$$p(y=1|x)=\sigma(0.3x_1+5.1x_2-0.5)$$

y: approval (1), rejection (0)
x_1: income
x_2: age

In this model, age is far more important than income! The older the applicant the more likely the approval.

Should we use this model to make real life loan decisions?

Quadratic Boundary

model for loan approval:

$$p(y=1|x)=\sigma(0.3x_1^2+5.1x_2^2-x_1x_2-0.5)$$

y: approval (1), rejection (0)
x_1: income
x_2: age

In this model, the effect of age and income are intertwined, $-x_1x_2$, and is much harder to interpret.

Should we use this model to make real life loan decisions?

EVALUATING CLASSIFIERS

EVALUATING LOGISTIC REGRESSION:

Once we have trained our logistic regression model, how do we evaluate model performance?

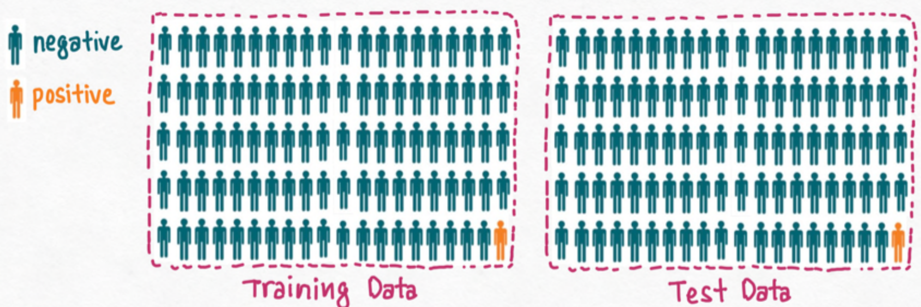
The easiest thing is to convert our soft decisions into hard decisions:

$$\hat{y} = \begin{cases} 1, & \sigma(f_w(x)) \geq 0.5 \\ 0, & \sigma(f_w(x)) < 0.5 \end{cases}$$

and then compute the accuracy of our classifier: $\dfrac{\#\text{ correctly classified}}{\text{total number of points}}$

Accuracy is intuitive to understand as an evaluation metric, but it can be extremely misleading!

Example: If your classifier for diagnosing a rare cancer is 99% accurate on training data and 99% accurate on test data, should we deploy this model to diagnose real patients?



If the disease is rare, only 1% of the population has it, then a classifier that classifies every patient as negative will be 99% accurate.

Data sets where one class out-numbers the other by a substantial amount is said to have class imbalance.

THE CONFUSION MATRIX:

Accuracy is potentially misleading as an evaluation metric because it averages the model performance on the majority class with its performance on the minority class. This allows good performance on the large class to obscure the poor performance on the small class.

We should examine how our predicted labels compare with the true labels in the data:

	Predicted $\hat{y}=1$	Predicted $\hat{y}=0$
Actual $y=1$	99	0
Actual $y=0$	1	0

correctly classified

incorrectly classified

This table is called the confusion matrix.