



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

Introduction to Logistic Regression

A Beginner Friendly Tutorial

Koushik Khan

✉ koushik.khan.ds@gmail.com [in linkedin.com/in/koushikkhan/](https://www.linkedin.com/in/koushikkhan/) [🔗 koushikkhan.github.io](https://github.com/koushikkhan)

March 3, 2025



Contents

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- 1 Motivation
- 2 ML for classification
- 3 Logistic regression
- 4 Parameter estimation
- 5 Model evaluation
- 6 Interpreting parameters
- 7 References
- 8 Thanks



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

MOTIVATION



Binary classification problem

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Imagine situations where we would like to know,
 - the *eligibility of getting a bank loan* given the value of credit score (x_{credit_score}) and monthly income (x_{income}).
 - identifying a *tumor as benign or malignant* given its size (x_{tumor_size}).
 - classifying an *email as promotional* given the no. of occurrences for some keywords like {'win', 'gift', 'discount'} ($x_{n_win}, x_{n_gift}, x_{n_discount}$).
 - finding a *monetary transaction as fraudulent* given the time of occurrence (x_{time_stamp}) and amount (x_{amount}).
- These problems occur frequently in real life & can be dealt with machine learning.
- All such problems come under the umbrella of what is known as *Classification*.
- In each scenario, only one of the two possible outcomes can occur, hence these are specifically known as *Binary Classification* problems.
- We need *special techniques* for such problems to handle with machine learning.



Datasets for binary classification problems

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- An ML algorithm needs supported data to learn patterns for addressing classification problems.
- Any dataset containing *numerical* or *categorical* features can be used for this.
- However, the *target variable must be categorical* in nature.
- Specifically, for binary classification, a target variable (Y) *must have two distinct values* like {'benign', 'malignant'}
- To use with ML algorithms, values of the target variable are *encoded into numeric representations* e.g. {"benign": 0, "malignant": 1}, a.k.a 'class-0' & 'class-1' respectively.



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

MACHINE LEARNING FOR CLASSIFICATION



How does machine learning help us for classification?

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- During inference, any ML model *predicts the class label* for an observation.
- A model estimates two probabilities \hat{p}_0 & \hat{p}_1 for the two classes where $\hat{p}_0 + \hat{p}_1 = 1$.
- The predicted label depends on $\max\{\hat{p}_0, \hat{p}_1\}$ i.e. it's the one which is most probable based on the given features.
- Sometimes, \hat{p}_1 is *compared with a predefined threshold* (that comes from experience!) p^* to take decision about the predicted class like below:

$$\begin{aligned}\text{predicted class} &= 1; \hat{p}_1 \geq p^* \\ &= 0; \text{otherwise}\end{aligned}$$



Why NOT linear regression?

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Can't we really use linear regression to address classification? The answer is *NO!*
- Let's try to understand why:
 - To estimate p_1 using linear regression, we need:

$$\hat{p}_1 = \alpha + \beta x_{\text{tumor_size}} \quad (1)$$

- Eqn. (1) *doesn't seem to be feasible* as the R.H.S, in principle, belongs to $(-\infty, +\infty)$ & the L.H.S belongs to $[0, 1]$.
- Can we convert $(\alpha + \beta x_{\text{tumor_size}})$ to something belonging to $[0, 1]$? That may work as an estimate of a probability! The answer is *YES!*
- We need a *converter (a function)*, say, $g(\cdot)$ that would map $(\alpha + \beta x_{\text{tumor_size}}) \in \mathbb{R}$ to some $p \in [0, 1]$.
- Fortunately, such functions do exist and they are often referred to as *link functions*.
- We are going to use a link function called *sigmoid function*, denoted by σ .



The sigmoid link function

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

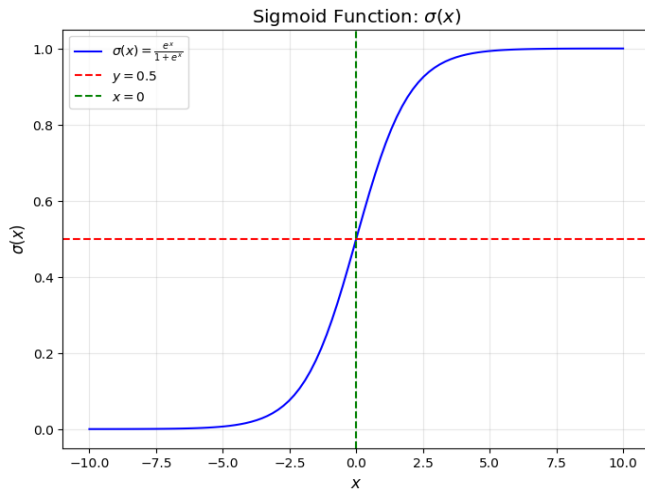


Figure: The sigmoid function



Math behind sigmoid

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks


- The *sigmoid function*, $\sigma(\cdot)$ is defined as below:

$$\sigma(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}; z \in \mathbb{R} \quad (2)$$

and the following is *always true*: $\sigma(z) \in [0, 1] \forall z \in \mathbb{R}$.

-  It has a very nice & important property:

$$\frac{d\sigma}{dz} = \sigma'(z) = \sigma(z) \times (1 - \sigma(z)) \text{ (verify!)} \quad (3)$$

-  It looks like an *elongated 'S'*, that is where it gets its name from.



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

LOGISTIC REGRESSION



The logistic regression setup

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- In logistic regression, we try to find $P(Y = 1)$ (i.e. *probability that a tumor will be malignant*) as below:

$$P(Y = 1) = p_1 = \frac{e^{\alpha + \beta x_{\text{tumor_size}}}}{1 + e^{\alpha + \beta x_{\text{tumor_size}}}} \quad (4)$$

- And this simplifies to:

$$\log\left(\frac{p_1}{1 - p_1}\right) = \alpha + \beta x_{\text{tumor_size}} \text{ (verify!)} \quad (5)$$

where the L.H.S is called *log-odds* or *logit*.



It's a linear classifier

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- The logistic regression equation (5) is actually a *straight line*.
- 🔄 **Recall** the prediction rule:

$$\begin{aligned}\text{predicted class} &= 1; p_1 \geq p^* \Rightarrow \alpha + \beta x_{\text{tumor_size}} \geq \log \left(\frac{p^*}{1 - p^*} \right) \\ &= 0; \text{otherwise}\end{aligned}$$

- A simple logistic regression (the one we discussed) predicts the class label by *identifying the regions on either side of a straight line* (or hyperplane in general), hence it's a *linear classifier*.



Linear versus non-linear classification

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

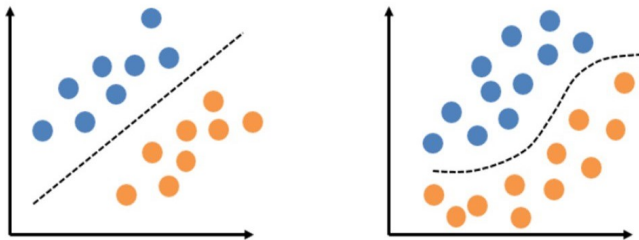


Figure: On the left, the classes are linearly separable, however they are not on the right

-  Logistic regression works well for *linearly separable* classes.



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

ESTIMATION OF PARAMETERS



Prerequisites

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters


References

Thanks

Bernoulli's Trial

A random experiment that results in one of the two possible outcomes, often called, a *success* and a *failure*, with a *constant probability* of success p .

Example

- tossing a fair coin - the coin shows either the 'HEAD' or the 'TAIL'
 - performing COVID test - result will be either '+ve' or '-ve'
 - detecting a tumor as 'benign' or 'malignant' etc.
-
- Bernoulli's trial can be expressed mathematically with a random variable, say, Y :
$$f(y|p) = P(Y = y) = p^y \times (1 - p)^{1-y}; y \in \{0, 1\} \quad (6)$$
where eqn. (6) is known as *Probability Mass Function (PMF)*.
 -  In PMF *we don't know* what value Y would take but we assume p *is known* to us.



Likelihood function

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- The PMF maps the values taken by Y to probabilities e.g.


$$\text{when } y = 1 \Rightarrow P(Y = 1) = p$$

$$\text{when } y = 0 \Rightarrow P(Y = 0) = (1 - p)$$

- Data involve N patients with: $\{(x_{i,tumor_size}, y_i)\}_{i=1}^N$, $y_i \in \{0, 1\}$.
- We consider Y_i takes the value y_i with probability p_i (*as each patient can have different probability to have a malignant tumor*).
- In reality, *we don't know these p_i 's*, so they are *estimated from data*.
- Interestingly, the expression for $P(Y_i = y_i) = p_i$ is same as before, eqn. (6),

$$f_i(p_i|y_i) = P(Y_i = y_i) = p_i^{y_i} \times (1 - p_i)^{1-y_i}; y_i \in \{0, 1\} \forall i = 1, 2, \dots, N \quad (7)$$

The eqn. (7) is known as the **likelihood** for Y_i taking a value y_i .

-  In Likelihood we know the data (*as we're observing them*), but the p_i 's are *unknown* to us.



Joint likelihood function

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks


- But what's the likelihood for *observing the entire dataset* i.e. $P(Y_1 = y_1 \cap \dots \cap Y_N = y_N)$? Well, *joint likelihood* gives that answer.
- It's computed as below:

$$L = P(Y_1 = y_1 \cap \dots \cap Y_N = y_N) = \prod_{i=1}^N p_i^{y_i} \times (1 - p_i)^{1-y_i} \quad (8)$$

Eqn. (8) is called the *joint likelihood* (L).

- It is much easier to work with joint likelihood after a *log-transformation*, also called the *log-likelihood* (LL),

$$LL = \log(L) = \sum_{i=1}^N \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} \quad (9)$$

-  Joint likelihood measures the probability of observing the underlying dataset i.e. having $\{Y_1 = y_1, \dots, Y_N = y_N\}$ for some unknown set of probabilities $\{p_1, \dots, p_N\}$.



Maximum likelihood estimation

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Let's assume, some process might have generated the observed dataset $\{(x_i, \text{tumor_size}, y_i)\}_{i=1}^N$
- But NOT sure what values of p_i 's the process would have considered to generate the dataset.
- We can imagine *several potential candidates* for each p_i (say, all belong to \mathbb{P}) that might have been used to generate the dataset.
- In principle, the best candidate for each p_i would be the one that makes the joint likelihood (L) or log likelihood (LL) (both are functions of p_i 's) *maximum*.
- Mathematically, we would use,

$$\arg \max_{p_1, \dots, p_N \in \mathbb{P}} \sum_{i=1}^N \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} \quad (10)$$

to find $\hat{p}_1, \dots, \hat{p}_N$.



Going deeper

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Finding $\hat{p}_1, \dots, \hat{p}_N$ using eqn. (10) is basically an *optimization problem*.
- 🔄 **Recall** $p_i = \frac{e^{\alpha + \beta x_{i,tumor_size}}}{1 + e^{\alpha + \beta x_{i,tumor_size}}} = \sigma(\alpha + \beta x_{i,tumor_size})$
- The eqn. (10) can be simplified as:

$$\arg \max_{p_1, \dots, p_N \in \mathbb{P}} \left[\sum_{i=1}^N \left\{ y_i \log(\sigma(\alpha + \beta x_{i,tumor_size})) + (1 - y_i) \log(1 - \sigma(\alpha + \beta x_{i,tumor_size})) \right\} \right] \quad (11)$$

- 📌 Here each p_i *is a function* of the parameters α & β and the known data $x_{i,tumor_size}$.
- 📌 Finding $\hat{p}_1, \dots, \hat{p}_N$ *is equivalent* to finding α and β with the help of eqn. (11).



Using gradient descent

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Specifically, it's a *maximization problem* here, however the ML community rather take *negative of LL*, *NLL* to *minimize* with the help of *Gradient Descent* !
- Actually a *scaled version* of NLL ($\frac{1}{N}NLL$) is used and that is known as *Binary Cross Entropy (BCE)* loss function.
- Now the problem boils down to:

$$\arg \min_{p_1, \dots, p_N \in \mathbb{P}} \frac{1}{N} NLL = \arg \min_{p_1, \dots, p_N \in \mathbb{P}} -\frac{1}{N} \sum_{i=1}^N \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} \quad (12)$$

or equivalently,

$$\arg \min_{\alpha, \beta} -\frac{1}{N} \sum_{i=1}^N \left\{ y_i \log(\sigma(\alpha + \beta x_{i, \text{tumor_size}})) + (1 - y_i) \log(1 - \sigma(\alpha + \beta x_{i, \text{tumor_size}})) \right\} \quad (13)$$



Computing derivatives

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Gradient Descent *computes derivatives* of BCE w.r.t α and β .

- Here is how it works:

Let us first consider $z_i = \alpha + \beta x_{i,tumor_size}$

we would like to compute:

$$\begin{aligned}\frac{\partial BCE}{\partial \alpha} &= -\frac{1}{N} \frac{\partial}{\partial \alpha} \sum_{i=1}^N \left\{ y_i \log(\sigma(z_i)) + (1 - y_i) \log(1 - \sigma(z_i)) \right\} \\ &= -\frac{1}{N} \sum_{i=1}^N \left\{ y_i \frac{\partial}{\partial \alpha} \log(\sigma(z_i)) + (1 - y_i) \frac{\partial}{\partial \alpha} \log(1 - \sigma(z_i)) \right\}\end{aligned}\quad (14)$$

and similarly,

$$\frac{\partial BCE}{\partial \beta} = -\frac{1}{N} \sum_{i=1}^N \left\{ y_i \frac{\partial}{\partial \beta} \log(\sigma(z_i)) + (1 - y_i) \frac{\partial}{\partial \beta} \log(1 - \sigma(z_i)) \right\} \quad (15)$$



The magic of using sigmoid

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- We will compute the derivatives one by one.
- Here is how we proceed for α :

$$\begin{aligned}\frac{\partial \log(\sigma(z_i))}{\partial \alpha} &= \overbrace{\frac{\partial \log(\sigma(z_i))}{\partial \sigma(z_i)} \times \frac{\partial \sigma(z_i)}{\partial z_i} \times \frac{\partial z_i}{\partial \alpha}}^{\text{chain rule of differentiation}} \\ &= \frac{1}{\sigma(z_i)} \times \sigma(z_i) \times (1 - \sigma(z_i)) \times 1 \\ &= (1 - \sigma(\alpha + \beta x_{i,tumor_size}))\end{aligned}\tag{16}$$

$$\begin{aligned}\frac{\partial \log(1 - \sigma(z_i))}{\partial \alpha} &= \frac{\partial \log(1 - \sigma(z_i))}{\partial (1 - \sigma(z_i))} \times \frac{\partial (1 - \sigma(z_i))}{\partial \sigma(z_i)} \times \frac{\partial \sigma(z_i)}{\partial z_i} \times \frac{\partial z_i}{\partial \alpha} \\ &= \frac{1}{(1 - \sigma(z_i))} \times (-1) \times \sigma(z_i) \times (1 - \sigma(z_i)) \times 1 \\ &= -\sigma(z_i) = -\sigma(\alpha + \beta x_{i,tumor_size})\end{aligned}\tag{17}$$



Continuing...

- And similarly for β :

$$\begin{aligned}\frac{\partial \log(\sigma(z_i))}{\partial \beta} &= \overbrace{\frac{\partial \log(\sigma(z_i))}{\partial \sigma(z_i)} \times \frac{\partial \sigma(z_i)}{\partial z_i} \times \frac{\partial z_i}{\partial \beta}}^{\text{chain rule of differentiation}} \\ &= \frac{1}{\sigma(z_i)} \times \sigma(z_i) \times (1 - \sigma(z_i)) \times x_{i,tumor_size} \\ &= (1 - \sigma(\alpha + \beta x_{i,tumor_size})) \times x_{i,tumor_size}\end{aligned}\tag{18}$$

$$\begin{aligned}\frac{\partial \log(1 - \sigma(z_i))}{\partial \beta} &= \frac{\partial \log(1 - \sigma(z_i))}{\partial (1 - \sigma(z_i))} \times \frac{\partial (1 - \sigma(z_i))}{\partial \sigma(z_i)} \times \frac{\partial \sigma(z_i)}{\partial z_i} \times \frac{\partial z_i}{\partial \beta} \\ &= \frac{1}{(1 - \sigma(z_i))} \times (-1) \times \sigma(z_i) \times (1 - \sigma(z_i)) \times x_{i,tumor_size} \\ &= -\sigma(\alpha + \beta x_{i,tumor_size}) \times x_{i,tumor_size}\end{aligned}\tag{19}$$

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks



The iterative rule of gradient descent

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- We can now compute $\frac{\partial BCE}{\partial \alpha}$ and $\frac{\partial BCE}{\partial \beta}$.
- And finally make use of the Gradient Decent *update rule*:

$$\begin{array}{c} \text{estimates at } (t+1)^{th} \text{ step} \\ \underbrace{\begin{pmatrix} \hat{\alpha}^{(t+1)} \\ \hat{\beta}^{(t+1)} \end{pmatrix}} \end{array} = \begin{array}{c} \text{estimates at } t^{th} \text{ step} \\ \underbrace{\begin{pmatrix} \hat{\alpha}^{(t)} \\ \hat{\beta}^{(t)} \end{pmatrix}} \end{array} - \eta \cdot \begin{array}{c} \text{gradient adjustments at } t^{th} \text{ step} \\ \overbrace{\begin{pmatrix} \left. \frac{\partial BCE}{\partial \alpha} \right|_{\hat{\alpha}^{(t)}} \\ \left. \frac{\partial BCE}{\partial \beta} \right|_{\hat{\beta}^{(t)}} \end{pmatrix}} \end{array} \quad (20)$$

Here η is the *learning rate*.



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

EVALUATING MODEL PERFORMANCE



Confusion matrix & related metrics

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation


Interpreting
parameters

References

Thanks

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure: Confusion Matrix

- All cases = $TP + TN + FP + FN$
- Correctly classified cases = $TP + TN$
- Misclassified cases = $FP + FN$
- Precision = $\frac{TP}{TP+FP}$
- Recall (Sensitivity) = $\frac{TP}{TP+FN}$
- Specificity = $\frac{TN}{TN+FP}$
- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
- F1-score = $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
-  All the above metrics except the first one *depends on the threshold p^** .



Understanding precision

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Precision (a.k.a **True Positive Rate**) measures the *probability of predicting a true positive case* by a fitted model.
- According to the formula: $\frac{TP}{TP+FP}$, the *lesser* the number of false positive cases, the *higher* will be the precision.
- Precision is important where is making *false positive mistakes* is *risky*.
- In a email spam detection system, it's crucial that a *non-spam* email is *not getting tagged* as a spam email, otherwise an user may miss an important email. - *expecting a high precision*
- In medical diagnosis (e.g. cancer detection) high precision gives *confidence* to the doctors to start treatment without further tests. - *expecting a high precision*
- Precision is also important when the dataset is highly imbalanced (e.g. credit fraud detection, where getting a fraudulent transaction is rare). If the precision is low, even if the accuracy is very high, the model would probably raise many *false alarms*, which is misleading. - *expecting a high precision*



Understanding recall

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- Recall (a.k.a **Sensitivity**) measures the *probability of detecting a true positive case* when it's *actually positive*.
- According to the formula: $\frac{TP}{TP+FN}$, the *lesser* the number of false negative cases, the *higher* will be the recall.
- Recall is important where making *false negative mistakes* is *risky*.
- In case of cancer detection diagnosis, it's very important that a cancer is *getting detected* in the body if it is *actually there*, otherwise it will be a *life risk*. - *expecting a high recall*
- While using medical kit for detecting COVID, it's important that a person is *NOT tagged* as "COVID -VE" when he is *actually* "COVID +VE". Having such cases will *infect* many other people. - *expecting a high recall*
- Detecting as many threats as possible is important for an airport security system. Having higher values of recall will make sure almost all positive cases are identified. - *expecting a high recall*



Understanding F1 score

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- The idea of using F1-score is to keep *a balance* between precision and recall.
- According to the formula: $\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$, F1-score calculates *harmonic mean* of precision and recall.
- Harmonic mean *penalizes the extreme values* of both precision and recall.
- For example, in case of credit fraud detection, actual fraudulent cases (positive) are very rare and this may make the model biased towards *legitimate cases* with a *very high accuracy*, however it *may not* make any sense.
- Here the model should actually:
 - have *high precision* i.e. lower chance of raising false alarm by identifying a legitimate case as fraudulent
 - have *high recall* i.e. lower chance of missing fraudulent transaction

The F1-score keeps a balance between these and gives a much more *realistic evaluation* of how well a model is performing in detecting the *minority class (fraud)*.



Receiver operating characteristic curve (ROC)

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- It's a *graphical tool* depending on two metrics derived from the *confusion matrix*:
 - True Positive Rate = $P(\text{Predicted Positive} \mid \text{Actually Positive}) = \frac{TP}{TP+FP}$
 - False Positive Rate = $P(\text{Predicted Positive} \mid \text{Actually Negative}) = \frac{FP}{FP+TN}$
- By *varying* p^* within a *permissible range* a set of $\{(FPR_k, TPR_k)\}_{k=1}^K$ are obtained, and are plotted to form what is called *ROC Curve*.
- The FPR varies *along the X-axis* and TPR varies *along the Y-axis*.
- Both TPR and FPR vary within $[0, 1]$ resulting the *total area* of the plotting canvas to be 1.
- The *diagonal line*, connecting the coordinates $(0, 0)$ and $(1, 1)$ indicates *$FPR = TPR$* , which is how a *random model* would behave.
- It's always good to have a model which produces *TPR values on the higher side* and *FPR values on the lower side*.



Area under the curve (AUC)

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- It's one of the robust measures to *compare* different models or model configurations.
- The diagonal line (---) divides the plotting canvas in to *two halves* having an *area of 0.5* each. This line indicates a random classifier which is *equally good and bad*.
- Any model better than the random one will *cover an area* > 0.5 .
- The *higher* the AUC the model achieves, the *better* its performance.

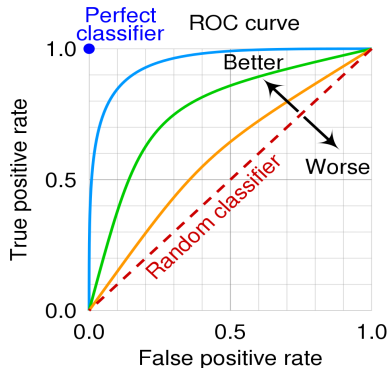


Figure: Comparing models with ROC curves and AUC values ¹

¹Image Source: [Wikipedia](#)



Interpretations of parameters

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression




Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

- α : When $x_{\text{tumor_size}}$ is zero, it's the *value of the log-odds*. It's often called the *baseline log-odds*.
- β : It's the *change in log-odds* for an *unit change in $x_{\text{tumor_size}}$* .
-  Interpretations are just like the *linear regression* as the alternate form of logistic regression (eqn. (5)) is exactly a linear regression w.r.t the log-odds.
-  $\log\text{-odds} \in (-\infty, +\infty)$ (*verify!*)
-  Even though logistic regression is used for *classification*, it actually *estimates a probability*, which is *continuous* within $[0, 1]$ - it's a *bridge* between *continuous modeling* and *discrete outcomes*, classification is just a *practical application* of this model.



References

Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression





Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

-  Generalized Linear Models by P. McCullagh, J. A. Nelder 
-  Categorical Data Analysis by Alan Agresti 



Logistic
Regression

Koushik
Khan

Motivation

ML for clas-
sification

Logistic
regression

Parameter
estimation

Model
evaluation

Interpreting
parameters

References

Thanks

Thank You