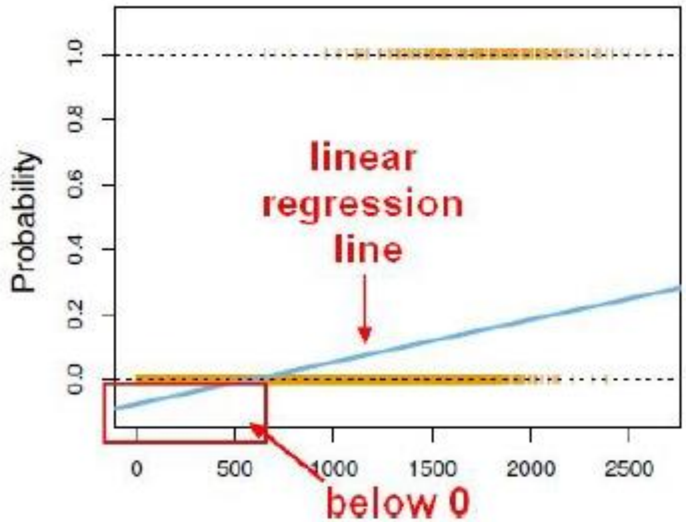
Introduction to Logistic Regression

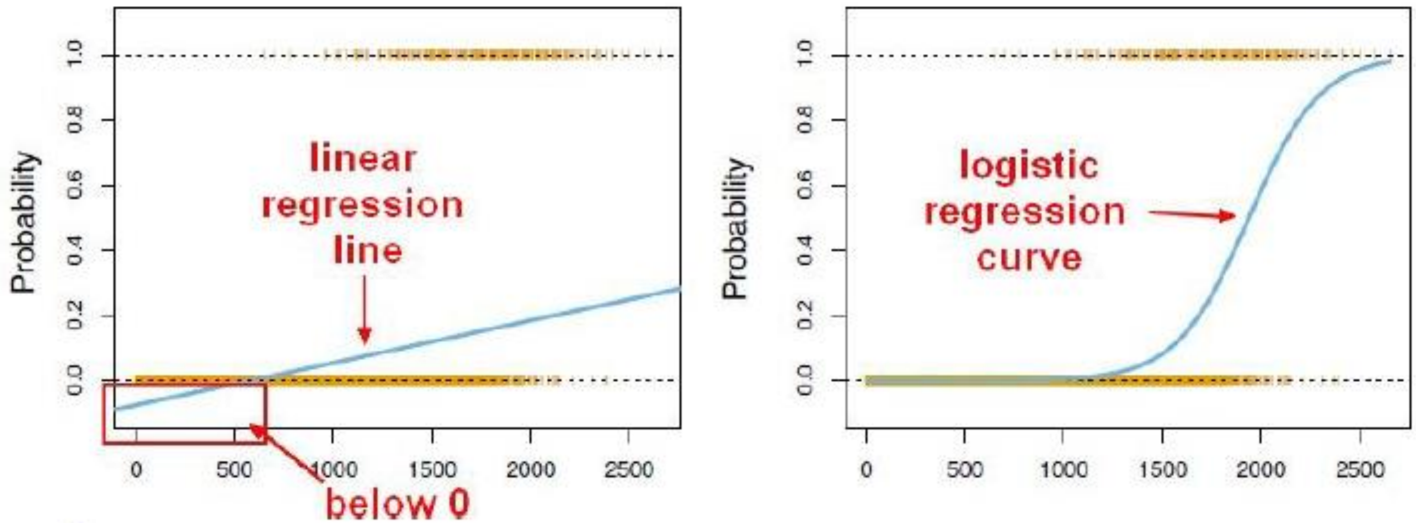
- We want to learn about Logistic Regression as a method for **Classification.**
- Some examples of classification problems:
  - Spam versus “Ham” emails
  - Loan Default (yes/no)
  - Disease Diagnosis
- Above were all examples of Binary Classification

- So far we’ve only seen regression problems where we try to predict a continuous value.
- Although the name may be confusing at first, logistic regression allows us to solve classification problems, where we are trying to predict discrete categories.
- The convention for binary classification is to have two classes 0 and 1.

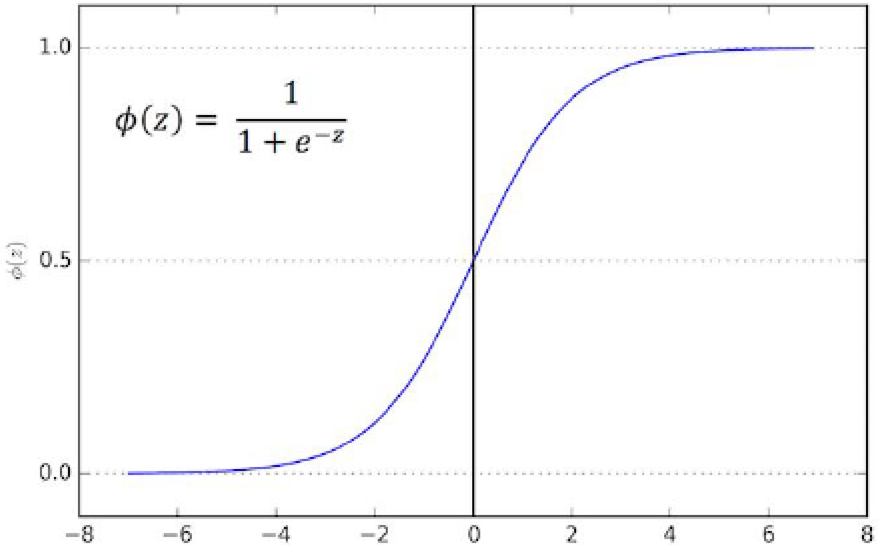
- We can't use a normal linear regression model on binary groups. It won't lead to a good fit:



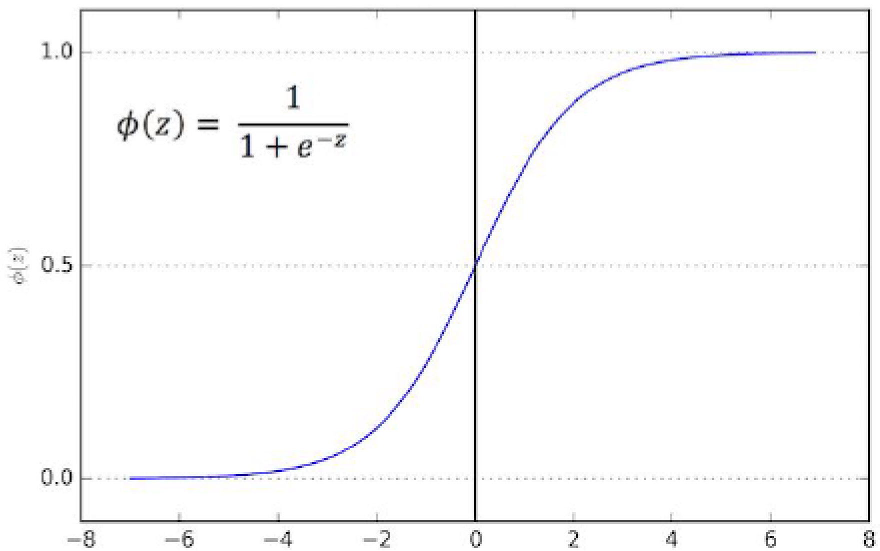
- Instead we can transform our linear regression to a logistic regression curve.



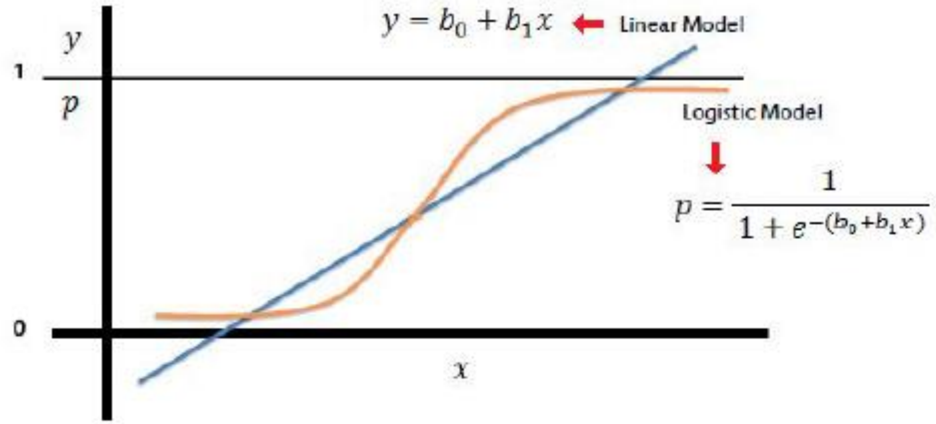
- The Sigmoid (aka Logistic) Function takes in any value and outputs it to be between 0 and 1.



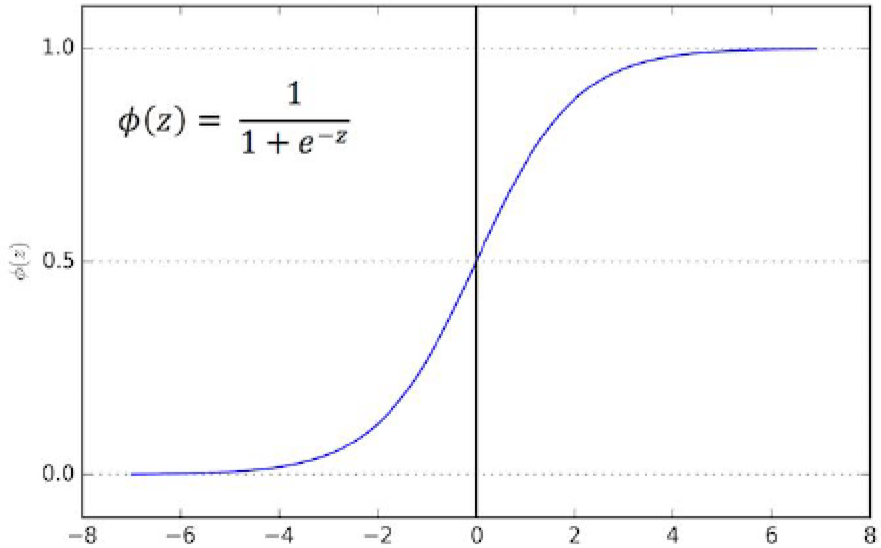
- This means we can take our Linear Regression Solution and place it into the Sigmoid Function.



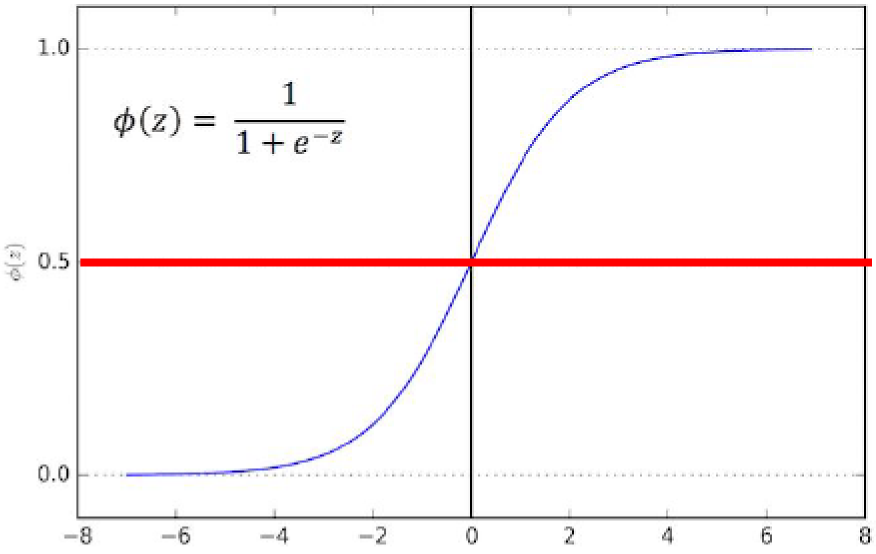
- This means we can take our Linear Regression Solution and place it into the Sigmoid Function.



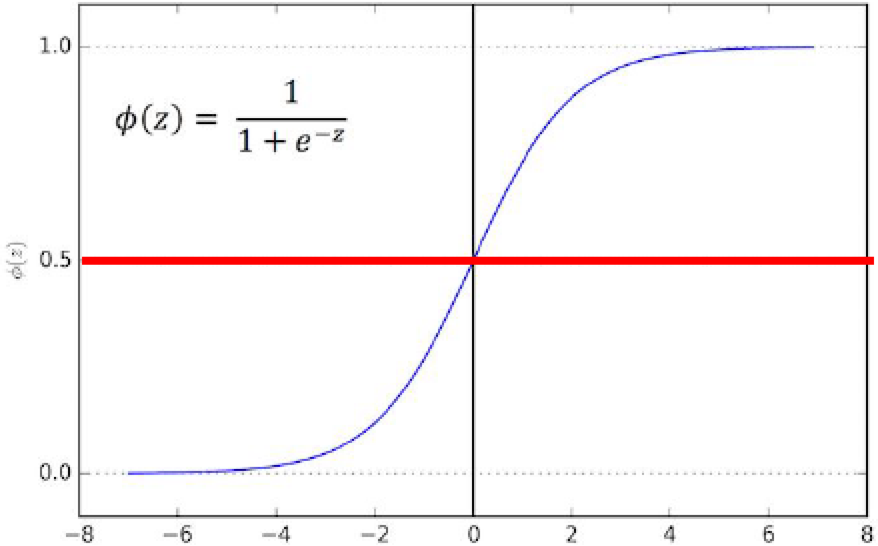
- This results in a probability from 0 to 1 of belonging in the 1 class.



- We can set a cutoff point at 0.5, anything below it results in class 0, anything above is class 1.



- We use the logistic function to output a value ranging from 0 to 1. Based off of this probability we assign a class.



- After you train a logistic regression model on some training data, you will evaluate your model's performance on some test data.
- You can use a confusion matrix to evaluate classification models.

- We can use a confusion matrix to evaluate our model.
- For example, imagine testing for disease.

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Example: Test for presence of disease
NO = negative test = False = 0
YES = positive test = True = 1

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Basic Terminology:

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Accuracy:

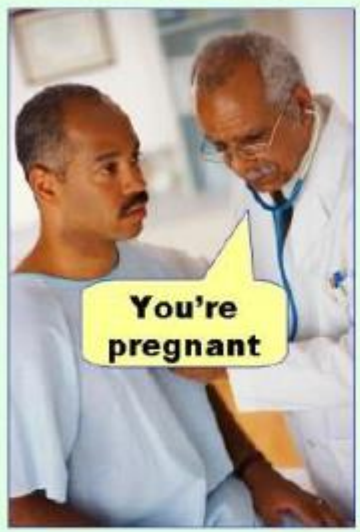
- Overall, how often is it **correct**?
- (TP + TN) / total = 150/165 = 0.91

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Misclassification Rate (Error Rate):

- Overall, how often is it **wrong**?
- (FP + FN) / total = 15/165 = 0.09

Type I error
(false positive)



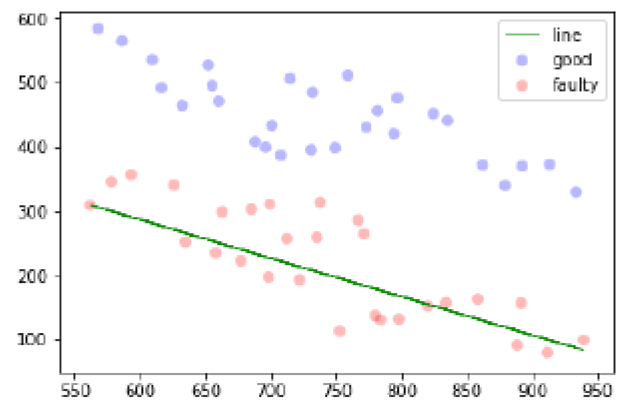
Type II error
(false negative)

Classification Problem

Linear Regression으로 학습해보자

$$f(x) = -2.4507508832597606 + 0.00227488 * RPM + 0.00379006 * Vibration$$

```
array([ 1.05856314e+00,  1.02370973e+00,  9.66120163e-01,
click to scroll output; double click to hide  7.49348117e-01,  1.03361935e+00,
        9.19162092e-01,  8.35784996e-01,  6.60707802e-01,
        6.44796277e-01,  6.26614080e-01,  7.88822723e-01,
        1.09355523e+00,  1.05263689e+00,  7.09256697e-01,
        7.61574640e-01,  1.21639013e+00,  9.41243609e-01,
        1.05646897e+00,  1.16639329e+00,  9.51115412e-01,
        1.13685352e+00,  1.12018650e+00,  9.20094079e-01,
        8.37485080e-01,  9.80760238e-01,  1.03990281e+00,
        9.22427788e-01, -1.14240215e-03,  1.75487796e-01,
        2.51301583e-01,  2.65731543e-01, -5.11098276e-02,
       -6.32186858e-02,  1.90719471e-01, -6.54767525e-02,
        2.55926977e-01, -1.16245894e-01,  3.18095711e-01,
        1.43005910e-01, -7.68091088e-02,  2.02908173e-01,
        4.18186047e-01, -3.09492679e-01,  3.78035795e-01,
        3.03754000e-01, -1.57109613e-01, -1.70750464e-01,
       -1.35112143e-01, -9.26369286e-03,  4.15348645e-02,
        1.18872240e-01, -8.57657338e-02,  1.67412730e-01,
       -7.89242969e-02,  6.05734081e-02])
```



Odds Ratio

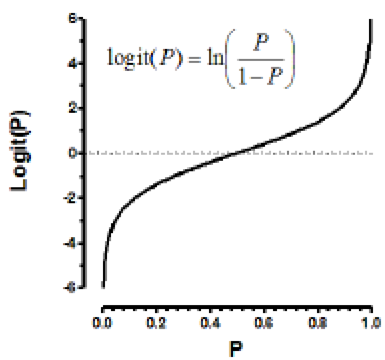
해당 사건이 일어날 확률과 일어나지 않을 확률의 비율

$$\frac{\text{일어날 확률}}{\text{일어나지 않을 확률}} \qquad \frac{P(X)}{1 - P(X)}$$

Logit function

X의 값이 주어졌을 때 y의 확률을 이용한 log odds

$$logit(p(y=1|x)) = log_e\left(\frac{p}{1-p}\right)$$



$$= log_e(p) - log_e(1-p)$$

$$= -\,log_e\left(\frac{1}{p} - 1\right)$$

Sigmoid(=Logistic) Function

Logit 함수의 역함수로 z에 관한 확률을 산출

$$f(z) = y = -log_e\left(\frac{1}{z} - 1\right)$$ 역함수로 바꾸면

$$z = -log_e\left(\frac{1}{y} - 1\right)$$ y에 관한 정리

Sigmoid(=Logistic) Function

$$z = -log_e\left(\frac{1}{y} - 1\right)$$

y에 관한 정리

$$e^{-z} = \frac{1-y}{y}$$

$$y * e^{-z} + y = 1$$

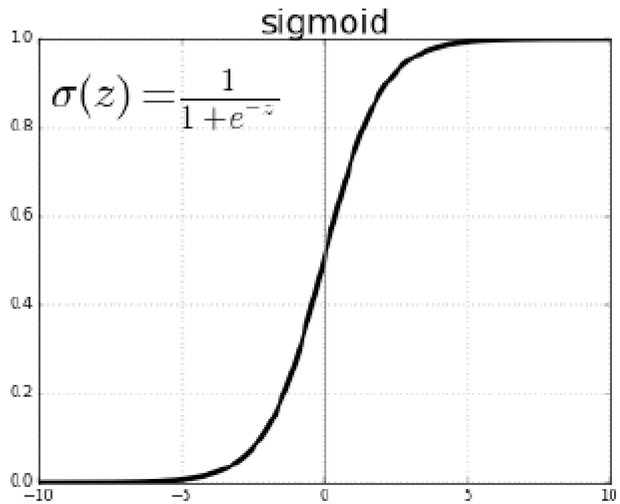
$$y(e^{-z} + 1) = 1$$

$$y = \frac{1}{1+e^{-z}}$$

Logistic Function =
Inverse of logit function

Sigmoid(=Logistic) Function

미분가능한 연속구간으로 변환
S형태로 닮았다고 하여 sigmoid function으로 호칭



Sigmoid(=Logistic) Function

선형함수에서 Sigmoid function으로 변환

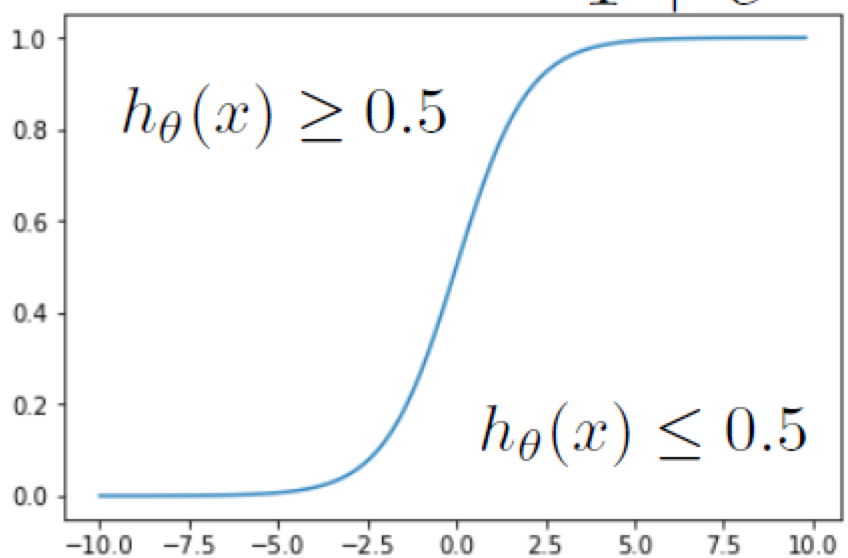
$$p = \sigma(z) = \frac{1}{1+e^{-z}}, \ \frac{p}{1-p} = \frac{\frac{1}{1+e^{-z}}}{\frac{e^{-z}}{1+e^{-z}}} = \frac{1}{e^{-z}} = e^{z}$$

$$log_e \frac{p}{1-p} = z$$

$$log_e \frac{p}{1-p} = z = w_0 x_0 + w_1 x_1 + \cdots + w_n x_n$$

가설 함수

$$h_\theta(x) = g(z) = \frac{1}{1+e^{-z}}$$



Training θ

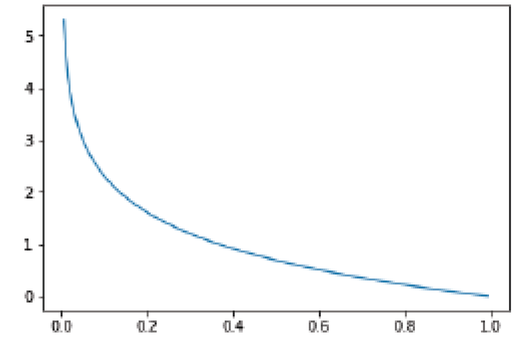
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T \mathbf{x}}}$$

ID	RPM	VIBRATION	STATUS
1	568	585	good
2	586	565	good
3	609	536	good
4	616	492	good
5	632	465	good
6	652	528	good
7	655	496	good
8	660	471	good
9	688	408	good
10	696	399	good

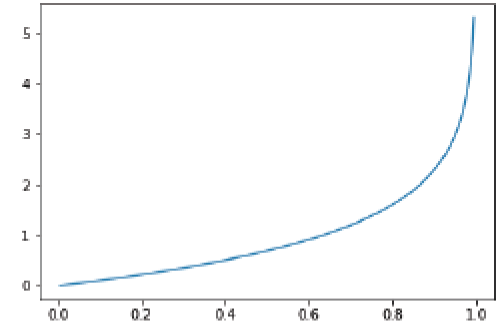
$$\theta^T \mathbf{x} = w_0 x_0 + w_1 x_1 + \cdots + w_n x_n$$

$$y = 0 \text{ or } 1$$

Cost Function



$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$



Partial derivation of cost function

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[-y^i(\log(1+e^{-\theta x^i})) + (1-y^i)(-\theta x^i - \log(1+e^{-\theta x^i}))\right]$$

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\theta x^i - \theta x^i - \log(1+e^{-\theta x^i})\right]$$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T \mathbf{x}}}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\theta x^i - \log(1+e^{\theta x^i})\right]$$

$$\begin{aligned} -\theta x^i - \log(1+e^{-\theta x^i}) &= -\left[\log e^{\theta x^i} + \log(1+e^{-\theta x^i})\right] \\ &= -\log(1+e^{\theta x^i}). \end{aligned}$$

Weight update

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i$$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

모든 $\boldsymbol{\theta_j}$ 동시에 업데이트

$$:= \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i$$

분류 문제의 정확도 성능

실제 Class 대비
얼마나 잘 맞혔는가?

Confusion Matrix (혼합 행렬)

- 실제 라벨과 예측 라벨의 일치 개수를 Matrix 형태로 표현하는 기법

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

Confusion Matrix (혼합 행렬)

True Positive (TP)

- 실제 결과 참(1)에 대한 예측이 맞음

True – 예측이 맞음

Positive – 참(1) 인 경우

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

Confusion Matrix (혼합 행렬)

True Negative (TN)

- 실제 결과 거짓(0)에 대한 예측이 맞음

True – 예측이 맞음

Negative – 거짓(0) 인 경우

		Prediction 1	Prediction 0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

Confusion Matrix (혼합 행렬)

False Positive (FP)

- 실제 결과 참(1)에 대한 예측이 틀림

False – 예측이 틀림

Positive – 참(1) 인 경우

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

Confusion Matrix (혼합 행렬)

False Negative (FN)

- 실제 결과 거짓(0)에 대한 예측이 틀림

False – 예측이 틀림

Negative – 거짓(0) 인 경우

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

Metrics for classification performance

- **Accuracy (정확도)** $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

- **Error Rate (오차율)** $Errorrate = \frac{FP+FN}{TP+TN+FP+FN} = (1 - Accuracy)$

- **Precision (정밀도)** $Precision = \frac{TP}{TP+FP}$ $(PPV: Positive\ Predict\ Value)$

- **Specificity (특이도)** $Specificity = \frac{TN}{TN+FP}$ $(TNR: True\ Negative\ Rate)$

- **Sensitivity (민감도)** $Sensitivity = \frac{TP}{TP+FP}$ $(TPR: True\ Positive\ Rate)$

정확도 (Accuracy, ACC)

- 전체 데이터 대비 정확하게 예측한 개수의 비율

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

$$ACC = 1 - ERR$$

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

오차율 (Error Rate, ERR)

- 전체 데이터 대비 부정확하게 예측한 개수의 비율

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

$$ERR = 1 - ACC$$

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

정밀도 (Precision, Positive Predictive Value)

- 긍정이라고 예측한 비율 중 진짜 긍정인 비율
- 긍정이라고 얼마나 잘 예측했는가? 긍정 예측 정밀도?

$$PRECISON(PPV) = \frac{TP}{TP + FP}$$

		Prediction 1	Prediction 0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

민감도 (Sensitivity, Recall, True Positive Rate)

- 실제 긍정 데이터중 긍정이라고 예측한 비율, 반환율, 재현율
- 얼마나 잘 긍정(예 - 암)이라고 예측하였는가?

$$RECALL(TPR) = \frac{TP}{TP + FN} = \frac{TP}{P}$$

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

특이성 (Specificity, True Negative Rate)

- 부정을 얼마나 잘 부정이라고 인식해는가?
- 전제 부정중 부정을 정확히 찾아낸 비율

$$SPC = \frac{TN}{TN + FP} = \frac{TN}{N}$$

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

F1 Score (F-measure, F-score)

- Precision과 Recall의 통합한 측정지표
- Precision과 Recall의 조화평균

$$F_1 = 2\frac{precision * recall}{precision + recall}$$

		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

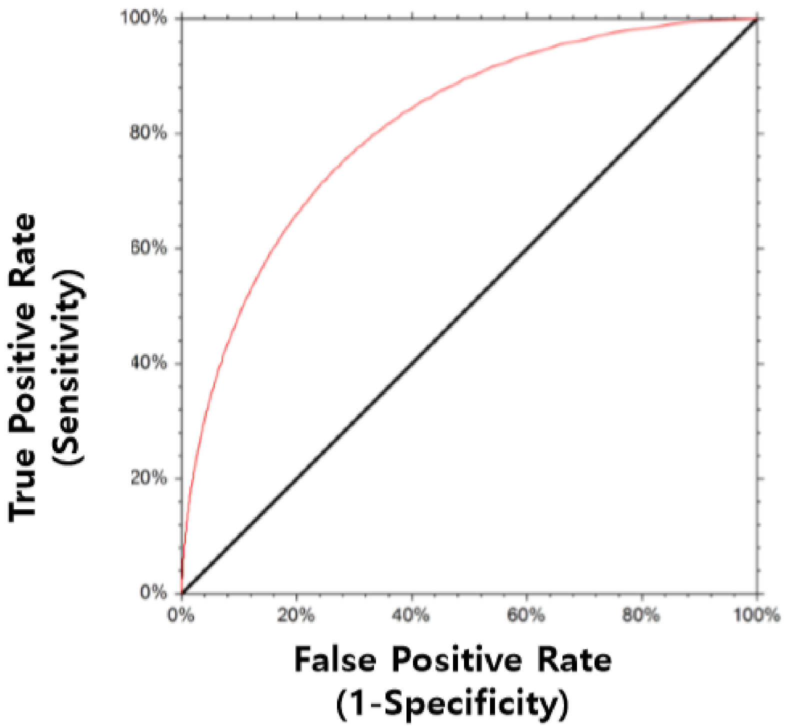
민감도-특이도
Trade-off?

Trade-Off 관계가 지표들 중 무엇을 선택해야 하나

ROC Curve

Receiver Operation Characteristics

Prediction Probability



Data	Class	Positive Prediction (Threshold)
1	P	0.9
2	P	0.8
3	N	0.7
4	P	0.6
5	P	0.55
6	N	0.54
7	N	0.53
8	N	0.51
9	P	0.5
10	N	0.4

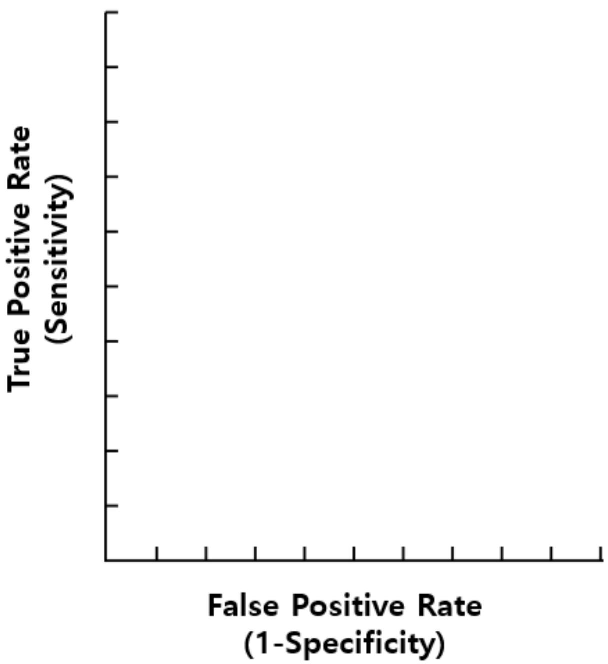
$$Sensitivity(TPR) = \frac{TP}{TP + FN} = \frac{TP}{P}$$

$$\mathrm{FPR} = 1 - Specificity(TNR)$$
$$= 1 - \frac{TN}{TN+FP} = 1 - \frac{TN}{N}$$

		Prediction	
		1	**0**
Actual Class	**1**	True Positive	False Negative
	0	False Positive	True Negative

Prediction Probability

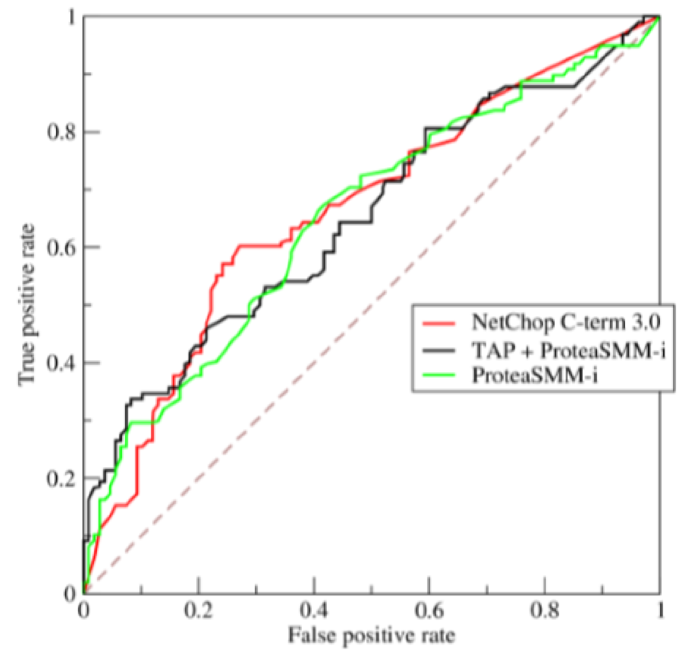
		Prediction	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative



Data	Class	Positive Prediction (Threshold)	TRUE Positive	FALSE Positive	TRUE Negative	FALSE Negative	TPR (Sensitivity)	FPR (1-Specificity)
1	P	0.9	1	0	5	4	0.2	0
2	P	0.8	2	0	5	3	0.4	0
3	N	0.7	2	1	4	3	0.4	0.2
4	P	0.6	3	1	4	2	0.6	0.2
5	P	0.55	4	1	4	1	0.8	0.2
6	N	0.54	4	2	3	1	0.8	0.4
7	N	0.53	4	3	2	1	0.8	0.6
8	N	0.51	4	4	1	1	0.8	0.8
9	P	0.5	5	4	0	1	1	1
10	N	0.4	5	5	0	0	1	1

AUC, Area Under Curve

- ROC curve의 하단의 넓이를 의미함
- ROC curve를 단순한 Single Metric (단 하나의 숫자)로 표현할 수 있음
- 대각선을 중심으로 상단에 붙어 있을 수록 높은 성능을 표시함



from Wikipedia(https://goo.gl/itMyAR)