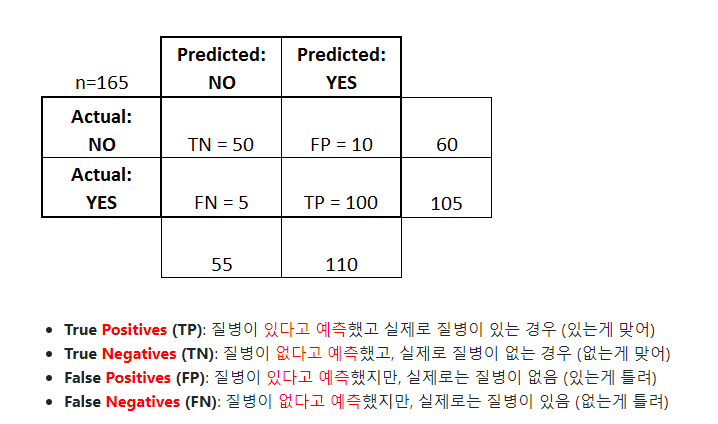
**혼동 행렬(confusion matrix)**

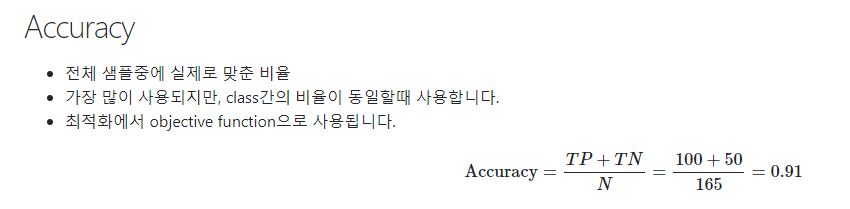
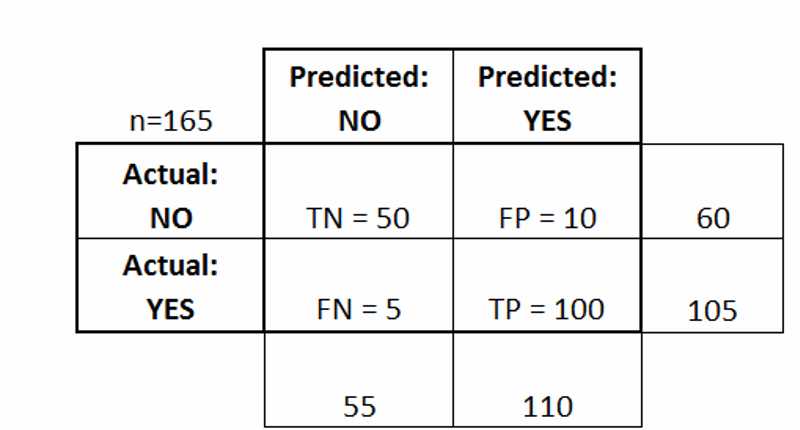
Ref : <http://incredible.ai/statistics/2018/06/30/Performance-Test/>

**분류모델**을 **평가** : 모델이 얼마나 정밀한지(잘~~ 맞추는지), 얼마나 **정확한 분류**를 했는지를 평가해야 한다.

**양성(Positive) : 1 , 암환자 🡨 극소수(학습이 잘 이루어지지 않는다)**음성(Negative) : 0, 정상인 ,

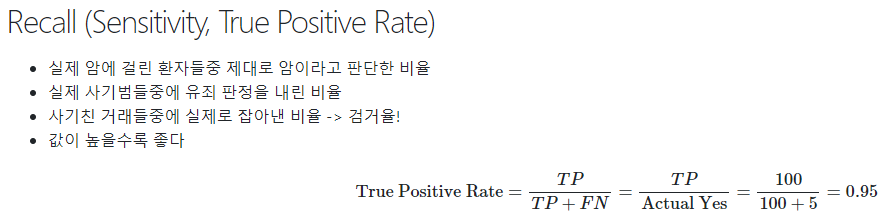


Ex) **암예측**모델 (정상인99명 암환자1명)  
accuracy(정확도) : 전체100명 전부 정상~~… 99점~~(상당히 높은 점수)==좋은모델?  
정상인99명이 중요한게 아니라 1명의 암환자를 맞춰야 한다.  
**양성(Positive)**을 얼마나 잘 맞췄냐 > 맞이 맞췄냐



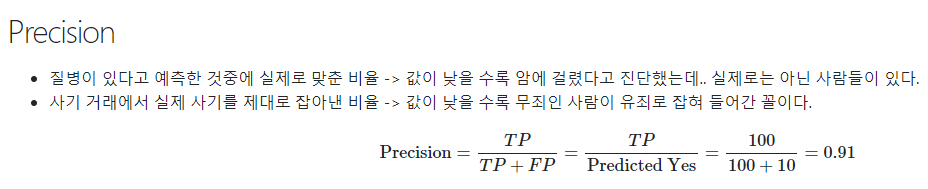
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* 재현률 : 실제 암(P)인데, 예측도 암(P) 비율 -----> **FN**
* **FN** : 실제 암(P)인데, 예측 정상(N) 🡪 **암환자 틀린것**

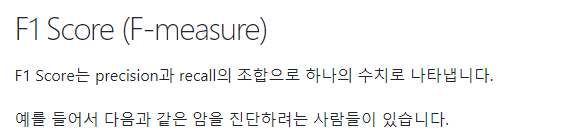


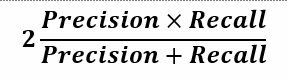
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

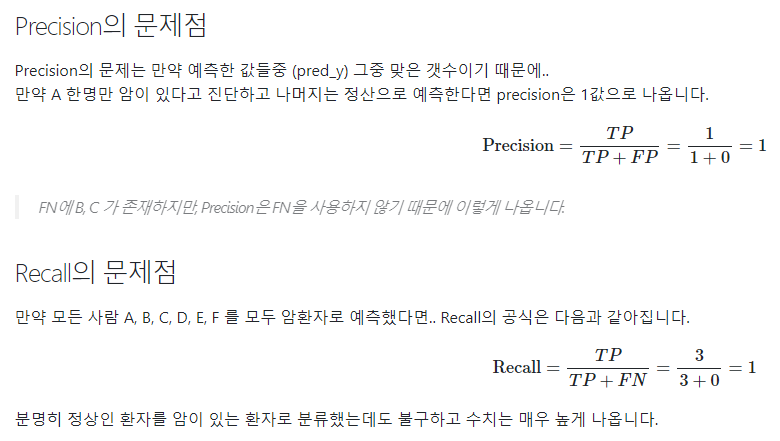
* ~~정밀도~~ : 예측도 암(P)인데 실제 암(P) 비율 -----> **FP**
* **FP**: 예측 암(P)인데 실제 정상(N) 🡪 **암 예측 틀린것**



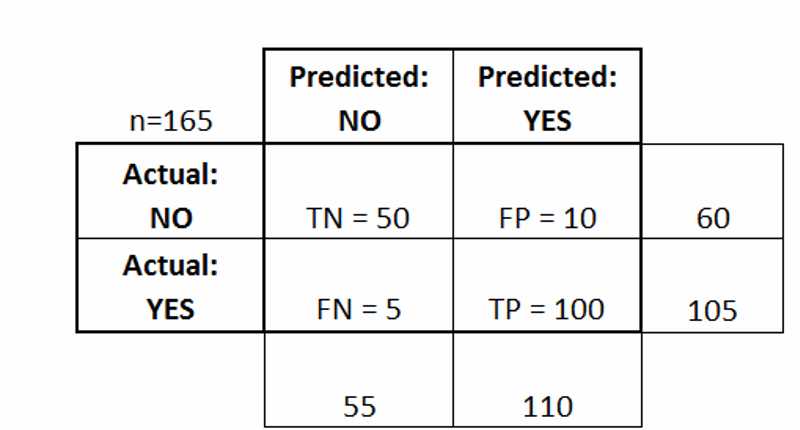
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_





🡪**F1점수를 본다**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

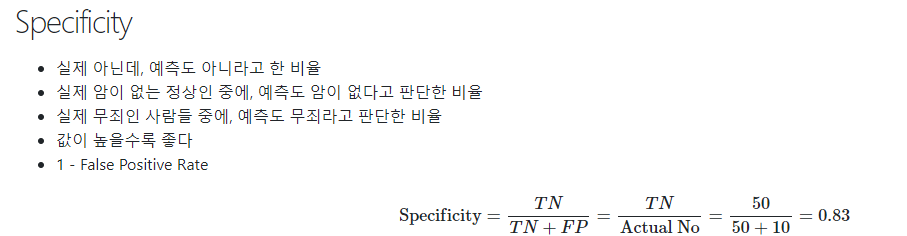


* **T**N : 정상환자 맞춘거  
  **T**P : 암환자 맞춘거  
  **FP**: 예측 암(P)인데 실제 정상(N) 🡪 **암 예측 틀린것**

**FN** : 실제 암(P)인데, 예측 정상(N) 🡪 **암환자 틀린것**

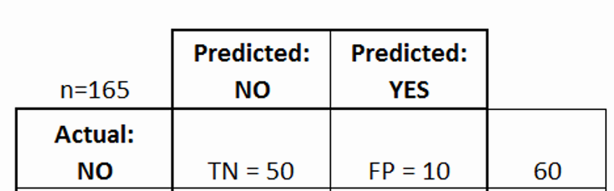
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* 정상인(N)을 기준으로 점수를 낸다 ……. XXXXXX
* ~~특이도~~ : 실제 정상(N)인데 예측 정상(N) 비율 -----> **TN**
* **TN**: 실제 정상(N)인데 예측 정상(N) 🡪 **정상인 잘 맞춘거**



**FPR**(False Positive Rate) = 🡨P

roc\_curv()에서 X축자리

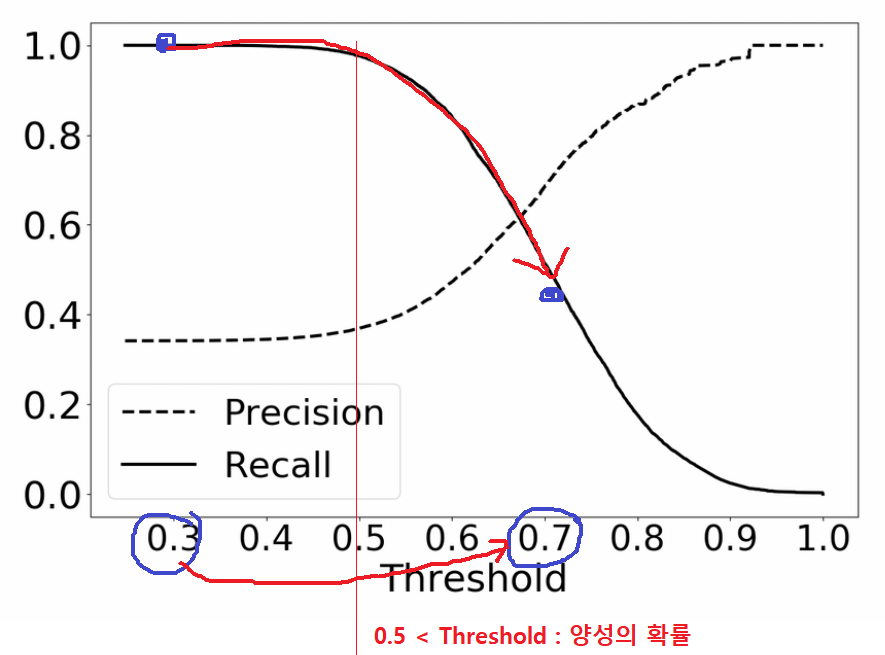


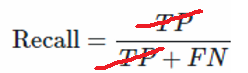
**정상인전체(60명, 비율1) = 특이도 + FPR**

**FPR = 1 - 특이도**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

precision\_recall\_curv란?

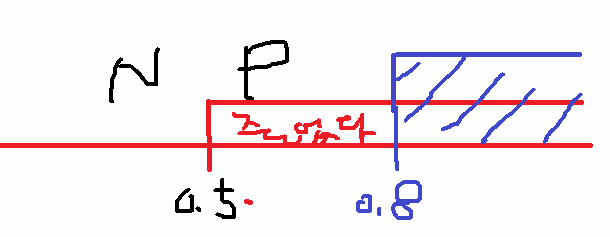
precision과 recall trade\_off관계 🡪 precision과 recall 변환에 따라 f1 점수가 달라진다.  
접점 근처에서 최적의 F1을 찾기 위함  


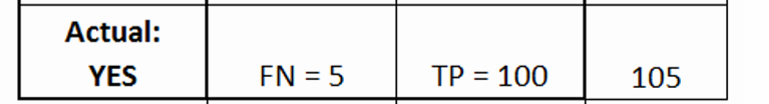
🡪 **FN** : 실제 암(P)인데, 예측 정상(N) 🡪 **암환자 틀린것**

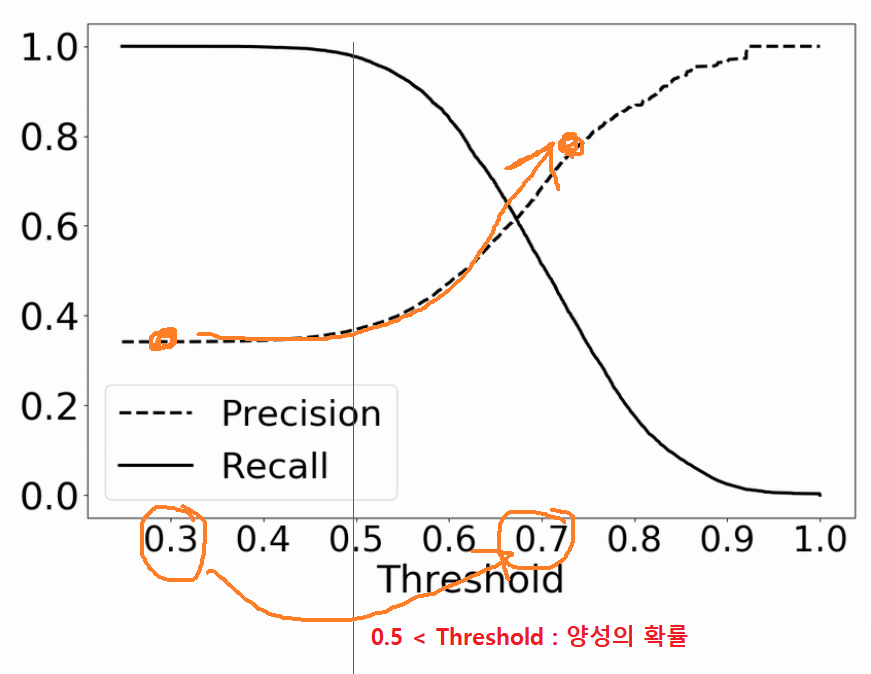
**FN20 🡪 FN40**

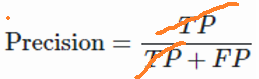
**FN이 커진다** == **암환자 틀린 것 수치가 커졌다.   
이전보다 많이 틀렸다 20명틀렸는데 이제는 40명 틀렸다  
암환자를 많이 못봐서 틀린 횟수가 늘었다**

* **FN이커진다 == 임계치가 커진다 == 암환자데이터가 줄었다**

****

****

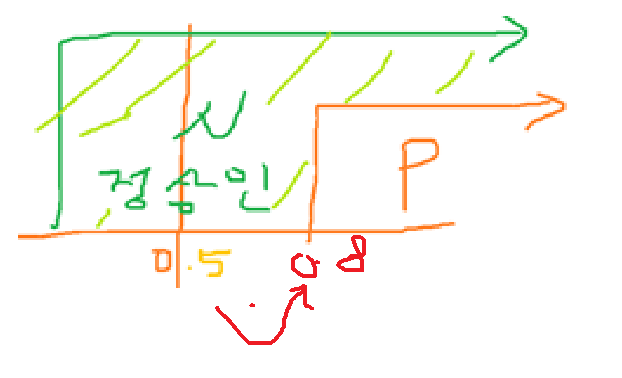
****

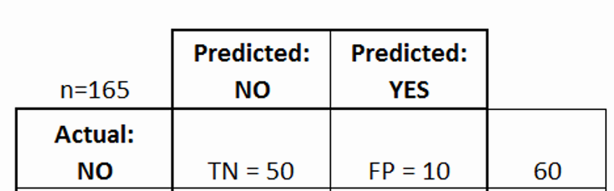
🡪 **FP**: 예측 암(P)인데 **실제 정상(N)** 🡪 **암 예측 틀린 것   
🡪 정상인 못맞추고 있는 상황**

**FP20 🡪 FP40**

**FP이 커진다** == **암환자라고 예측을 틀리게 한 수치가 커졌다.   
이전보다 많이 예측을 잘못했다 🡪 20명 잘못예측 이제는 40명 잘못예측  
정상인을 많이 못봐서 예측 틀린 횟수가 늘었다**

* **FP커진다 == 정상인을 잘 못맞춘다 == 임계치가 커진다 == 정상인 데이터가 늘었다**

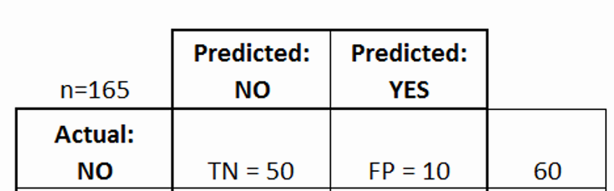
****

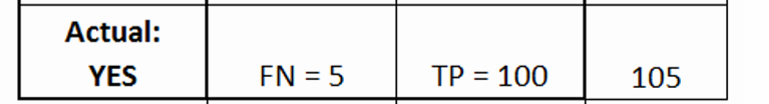


**임계값 변경에 따른 정밀도와 재현율 변화관계**

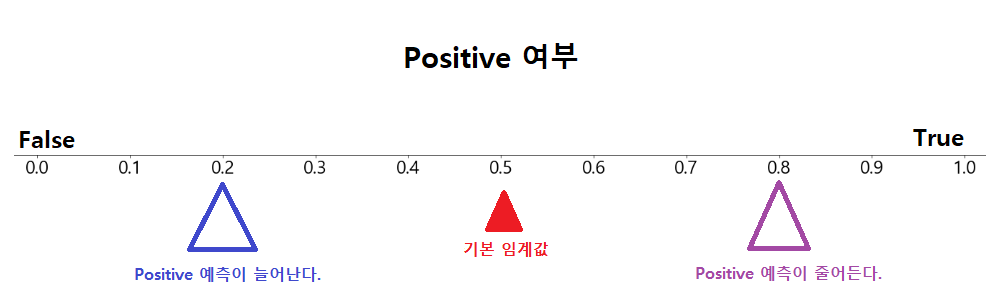
Ref : <https://ysyblog.tistory.com/72>

* 임계값을 높이면 **음성(N=정상인)으로 예측되는 샘플이 많아 진다.** 그래서 **정밀도(FP=음성)는 높아지고** 🡨🡪 **재현율(FN=양성)은 낮아진다.**

**정상인=N=음성** === FP : 정밀도

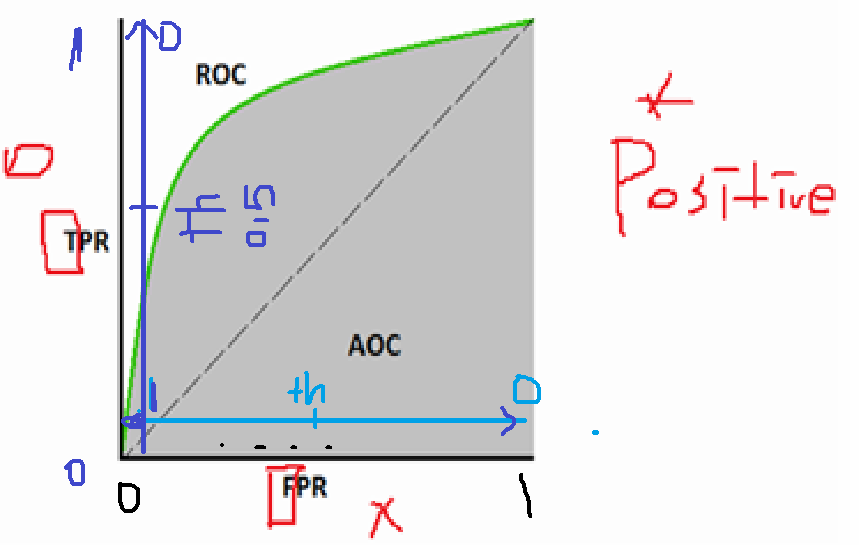
**암환자=P=양성**=== FN : 재현율

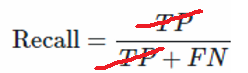
* 임계값을 낮추면 **양성(P=암환자)으로 예측되는 샘플이 많아 진다.**그래서 **정밀도(FP=음성)는 낮아지고** 🡨🡪 **재현율(FN=양성)은 높아진다.**
* 임계값을 변화시켰을때 **재현율과 정밀도는 음의 상관관계를 가진다.**

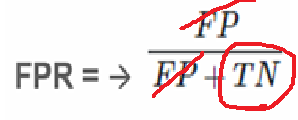


ROC 커브곡선이란?

위함양성(P=암환자) 맞췄다 : T**PR** 틀렸다(F**PR** == (1-특이도))

****

**TPR = 민감도(sensitivity) = 재현율(recall) vs. FPRTPR =** 🡪 **FN :   
TPR이 커진다 == FN이 작아진다 == 임계치가 작아진다**

* **[recall에 있던 말] FN이커진다 == 임계치가 커진다 == 암환자데이터가 줄었다  
  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
   🡪 TN**

**FPR이 커진다 == TN이 작아진다(정상인을 잘 못맞춘다=정상인데이터가 줄어든다) = 임계치가 작아진다.**

* **TN**: 실제 정상(N)인데 예측 정상(N) 🡪 **정상인 잘 맞춘거**