Review

Correction and Supplement

Corrections

Book and slide

- Book (Ch. 7, pp 204) 책이잘못됨. ppt는 수정X
  - Specificity = TN/(FP + TN) = True negative rate = 1 – False positive rate
  - Sensitivity = TP/(TP + FN) = True positive rate

- Slide (Ch. 7, pp 29)

$$\begin{aligned}\text{expected profit} = \; & p(\mathbf{p}) \cdot \boxed{[p(\mathbf{Y} \mid \mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot c(\mathbf{N}, \mathbf{p})]} \; + \\ & p(\mathbf{n}) \cdot \boxed{[p(\mathbf{N} \mid \mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y} \mid \mathbf{n}) \cdot c(\mathbf{Y}, \mathbf{n})]}\end{aligned}$$

Text book (pp. 204)

+ *FP*), which is the accuracy over the cases predicted to be positive. The *F-measure* is the harmonic mean of precision and recall at a given point, and is:

$$\text{F-measure} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Practitioners in many fields such as statistics, pattern recognition, and epidemiology speak of the sensitivity and specificity of a classifier:

$$\text{Sensitivity} = TN \,/\, (TN + FP) = \text{True negative rate} = 1 - \text{False positive rate}$$

$$\text{Specificity} = TP \,/\, (TP + FN) = \text{True positive rate}$$

You may also hear about the *positive predictive value*, which is the same as precision.

Accuracy, as mentioned before, is simply the count of correct decisions divided by the total number of decisions, or:

$$Accuracy = \frac{TP + TN}{P + N}$$

Swets (1996) lists many other evaluation metrics and their relationships to the confusion matrix.

Slide (Ch. 7, pp. 29)

- We now can deal with our motivating example
  - Instead of computing accuracies for the competing model, we would compute ***expected values***

- Furthermore, we can compare the two models easily for ***various*** distributions
  - For each distribution, we can simply replace the ***priors***
    - (ex) A unbalanced distribution: $p(\mathbf{p}) = 0.7, p(\mathbf{n}) = 0.3$
    - (ex) A balanced distribution: $p(\mathbf{p}) = 0.5, p(\mathbf{n}) = 0.5$

$$\begin{aligned}\text{expected profit} = \; & p(\mathbf{p}) \cdot [p(\mathbf{Y} \mid \mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot c(\mathbf{N}, \mathbf{p})] + \\ & p(\mathbf{n}) \cdot [p(\mathbf{N} \mid \mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y} \mid \mathbf{n}) \cdot c(\mathbf{Y}, \mathbf{n})]\end{aligned}$$

The other factors in the equation will ***not*** change

Supplements

Ch.9: Advantages of Naive Bayes (1/2)

- It is a very ***simple*** classifier
  - Yet it still takes all the feature evidence into account
- It is very ***efficient*** in terms of storage space and execution time
  - **Training**: consists only of storing $p(c)$ and $p(e_i \mid c)$ for each c and e_i
    - $p(c)$: we count the proportions of examples of class c among all examples
    - $p(e_i \mid c)$: we count the proportion of examples in class c for which e_i appears
  - **Classification**: requires only simple multiplications of them
- In spite of its simplicity and the strict independence assumption, it performs ***surprisingly well*** on many real-world tasks
  - Because the violation of the independence assumption tends not to hurt classification performance
  - What if two pieces of evidence are actually NOT independent and we treat them as being independent? ➔double counting of the evidence
    - However, double counting will not tend to hurt us (i.e., probability will be simply overestimated)
    - E.g.) P(AB) = P(A) x P(B|A) *vs.* P(AB) = P(A) x P(B)

Ch. 7: Alternative Calculation of EV (4/4)

- We now can deal with our motivating example
  - Instead of computing accuracies for the competing model, we would compute ***expected values***

- Furthermore, we can compare the two models easily for ***various*** distributions
  - For each distribution, we can simply replace the ***priors***
    - (ex) A unbalanced distribution: $p(\mathbf{p}) = 0.7$, $p(\mathbf{n}) = 0.3$
    - (ex) A balanced distribution: $p(\mathbf{p}) = 0.5$, $p(\mathbf{n}) = 0.5$

$$\begin{aligned}\text{expected profit} = \; & p(\mathbf{p}) \cdot [p(\mathbf{Y} \mid \mathbf{p}) \cdot b(\mathbf{Y}, \mathbf{p}) + p(\mathbf{N} \mid \mathbf{p}) \cdot c(\mathbf{N}, \mathbf{p})] + \\ & p(\mathbf{n}) \cdot [p(\mathbf{N} \mid \mathbf{n}) \cdot b(\mathbf{N}, \mathbf{n}) + p(\mathbf{Y} \mid \mathbf{n}) \cdot c(\mathbf{Y}, \mathbf{n})]\end{aligned}$$

The other factors in the equation will ***not*** change

Example

Model A

	p	n
Y	500	200
N	0	300

p:n = 5:5

P(Y|p) = 500/500 = 1
P(N|p) = 0/500 = 0

P(Y|n) = 200/500 = 0.4
P(N|n) = 300/500 = 0.6

Idea

	p	n
Y	100	0
N	0	900

p:n = 1:9

➔ Accuracy 100%

	p	n
Y	TP rate P(Y|p)	FP rate P(Y|n)
N	FN rate P(N|p)	TN rate P(N|n)

Model A

	p	n
Y	**100**	**360**
N	**0**	**540**

p:n = 1:9

P(Y|p) = 100/100 = 1
P(N|p) = 0/100 = 0

P(Y|n) = 360/900 = 0.4
P(N|n) = 540/900 = 0.6

$$\text{expected profit} = p(\mathbf{p})\cdot[p(\mathbf{Y}\mid\mathbf{p})\cdot b(\mathbf{Y},\mathbf{p}) + p(\mathbf{N}\mid\mathbf{p})\cdot c(\mathbf{N},\mathbf{p})] + p(\mathbf{n})\cdot[p(\mathbf{N}\mid\mathbf{n})\cdot b(\mathbf{N},\mathbf{n}) + p(\mathbf{Y}\mid\mathbf{n})\cdot c(\mathbf{Y},\mathbf{n})]$$

Ch. 7: Problems with Unbalanced Classes (2/3)

- Even when the skew is not so great, accuracy can be greatly ***misleading***

- Example: consider again our cellular-churn problem
  - Suppose A and B build their own churn prediction models
  - In a test set of 1,000 customers, the confusion matrices are as follows

A's model

	churn	not churn
Y	500	200
N	0	300

accuracy는 같지만

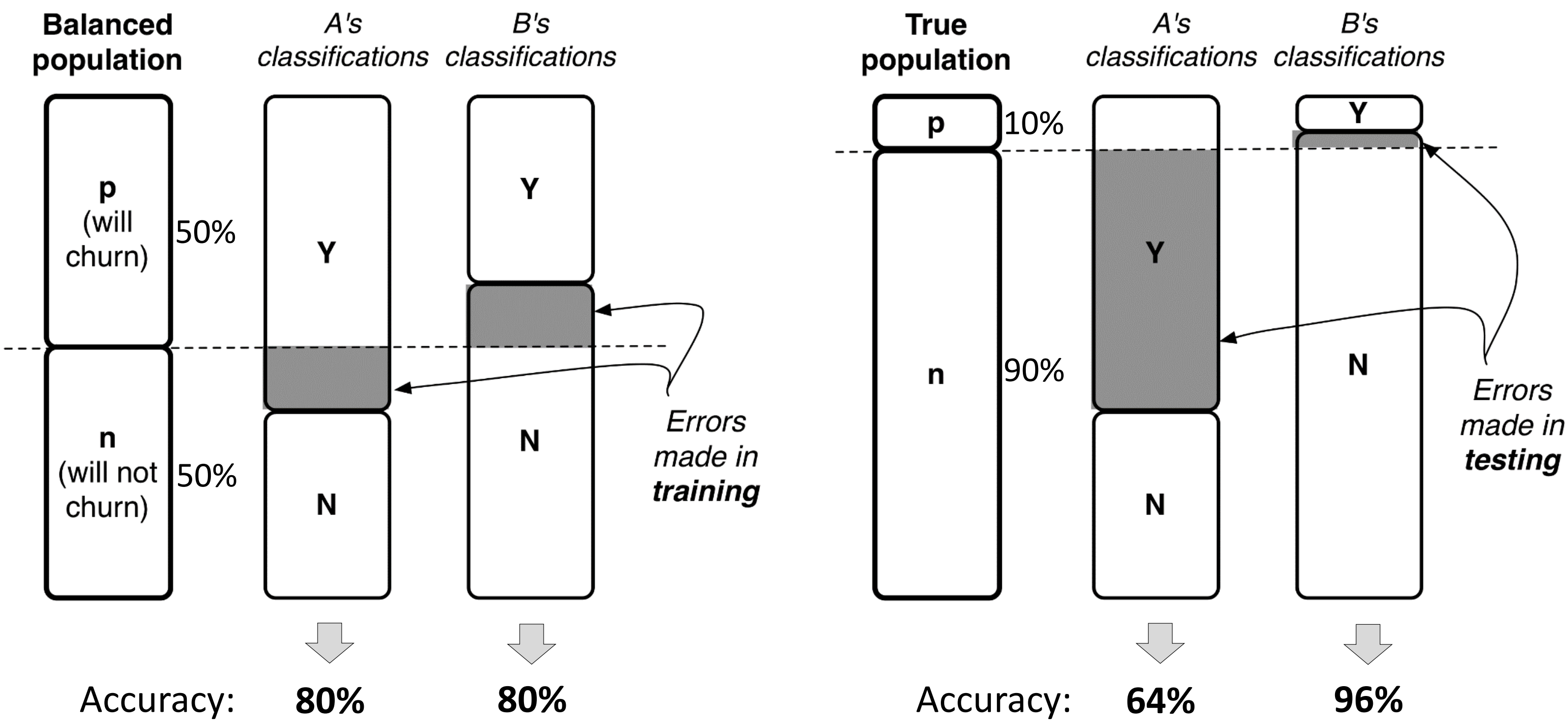
	churn	not churn
Y	300	0
N	200	500

B's model

    - A's model: correctly classifies 100% of "churn" but only 60% of "not churn"
    - B's model: correctly classifies 100% of "not churn" but only 60% of "churn"
  - Though they operate very differently, their accuracy are the ***same*** as 80%

Ch. 7: Problems with Unbalanced Classes (3/3)

- Furthermore, their accuracy changes with different test sets



Model B is better than model A ?

Example

	p	n
Y	TP rate P(Y|p)	FP rate P(Y|n)
N	FN rate P(N|p)	TN rate P(N|n)

Model **A**
(Accuracy 80%)

	p	n
Y	**500**	200
N	0	**300**

p:n = 5:5

TPR = 500/500 = 1
FNR = 0/500 = 0

FPR = 200/500 = 0.4
TNR = 300/500 = 0.6

Model **B**
(Accuracy 80%)

	p	n
Y	**300**	0
N	200	**500**

p:n = 5:5

TPR = 300/500 = 0.6
FNR = 200/500 = 0.4

FPR = 0/500 = 0
TNR = 500/500 = 1

Ideal
(Accuracy 100%)

	p	n
Y	100	0
N	0	900

p:n = 1:9

➔ Accuracy 100%

Model **A**

	p	n
Y	**100**	360
N	0	**540**

p:n = 1:9

TPR = 100/100 = 1
FNR = 0/100 = 0

FPR = 360/900 = 0.4
TNR = 540/900 = 0.6

Accuracy: 64% = 640/1000

Model **B**

	p	n
Y	**60**	0
N	40	**900**

p:n = 1:9

TPR = 60/100 = 0.6
FNR = 40/100 = 0.4

FPR = 0/900 = 0
TNR = 900/900 = 1

Accuracy: 96% = 960/1000