

## 베이지안(Baysian)

<http://contents.kocw.net/KOCW/document/2014/deagucatholic/ohkwangsik/8.pdf>

베이지안 확률 모델은 주관적인 추론을 바탕으로 만들어진 '사전확률'을 추가적인 관찰을 통한 '사후확률'로 업데이트하여 불확실성을 제거할 수 있다고 믿는 방법.

베이즈 정리는 posterior 확률을 찾는 과정이고 베이즈 추론을 MAP(Maximum a Posteriori)문제라고 부르기도 한다.

실습.

=====

```
install.packages("e1071")
```

```
install.packages("caret")
```

```
library(e1071)
```

```
data <- read.csv(file = "heart.csv", header = T)
```

```
head(data)
```

```
str(data)
```

```
library(caret)
```

```
set.seed(1234)
```

```
tr_data <- createDataPartition(y=data$AHD, p=0.7, list=FALSE)
```

```
tr <- data[tr_data,]
```

```
te <- data[-tr_data,]
```

```
Bayes <- naiveBayes(AHD~. ,data=tr)
```

```
Bayes
```

```
predicted <- predict(Bayes, te, type="class")
```

```
table(predicted, te$AHD)
```

```
AHD <- as.factor(te$AHD)
```

```
confusionMatrix(predicted, AHD)
```

```
=====
```

중간 결과 생략

```
>
> predicted <- predict(Bayes, te, type="class")
> table(predicted, te$AHD)
```

```
predicted No Yes
      No  45  12
      Yes   4  29
```

```
>
> AHD <- as.factor(te$AHD)
> confusionMatrix(predicted, AHD)
Confusion Matrix and Statistics
```

```
              Reference
Prediction No Yes
      No  45  12
      Yes   4  29
```

```
              Accuracy : 0.8222
              95% CI   : (0.7274, 0.8948)
No Information Rate : 0.5444
P-Value [Acc > NIR] : 2.84e-08
```

```
              Kappa : 0.6358
```

```
McNemar's Test P-Value : 0.08012
```

```
              Sensitivity : 0.9184
              Specificity : 0.7073
              Pos Pred Value : 0.7895
              Neg Pred Value : 0.8788
              Prevalence : 0.5444
              Detection Rate : 0.5000
              Detection Prevalence : 0.6333
              Balanced Accuracy : 0.8128
```

```
              'Positive' Class : No
```

```
> |
```