Introduction to Machine Learning

劉昭麟

國立政治大學資訊科學系

Categorical Classification

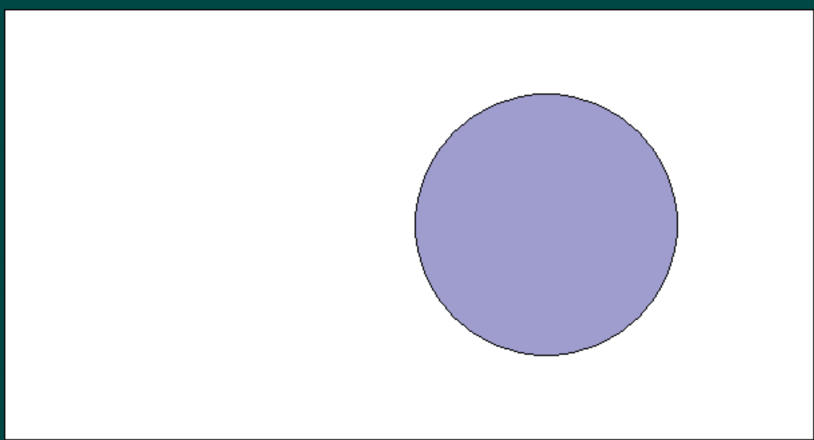
開學時就提到過的

- Training data and test data
  - Check Moodle
  - Deliberation
    - How do we split data into training and test?
- Validation
- DM4 Sections 5.1, 5.3, 5.4
- DM4 Section 5.8
  - Two classes classification
    - TP, FP, TN, FN

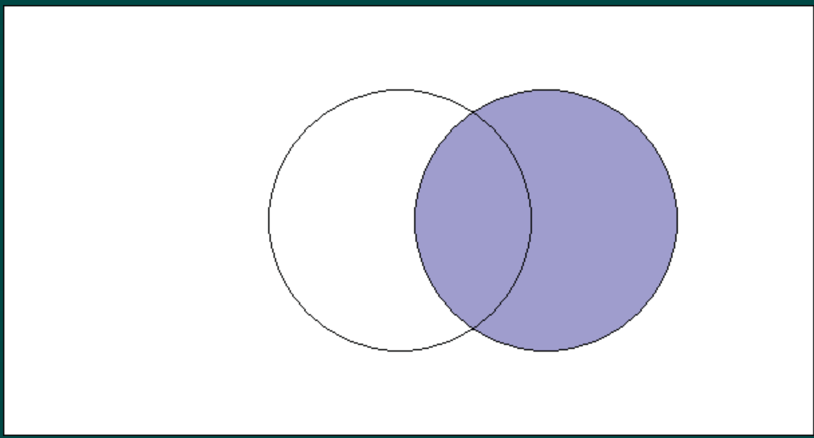
PRECISION, RECALL, AND F

Precision, Recall, and F Measures

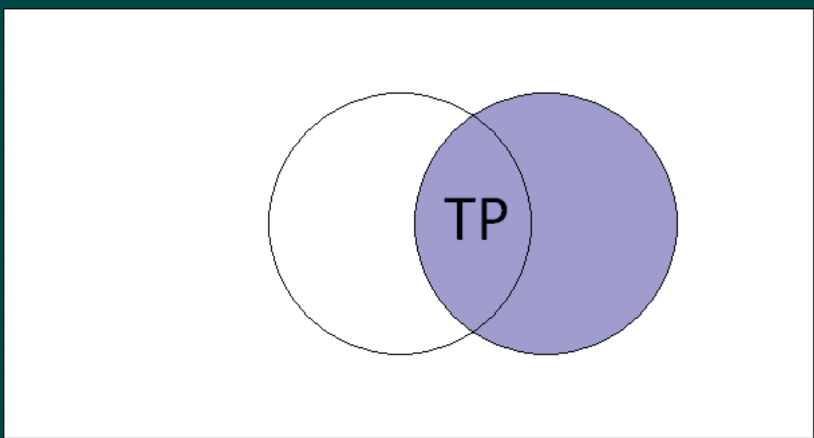
Precision, Recall, and F Measures



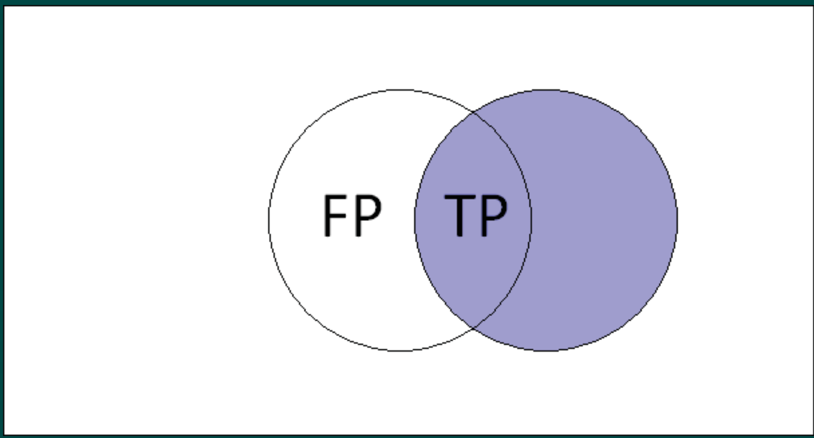
Precision, Recall, and F Measures



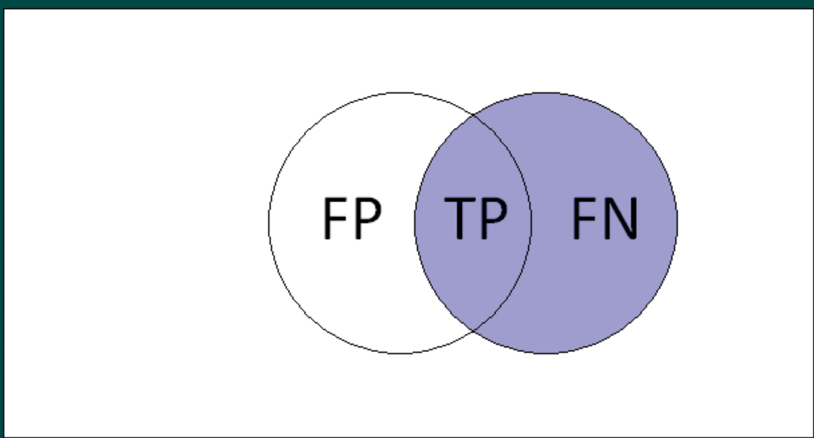
Precision, Recall, and F Measures



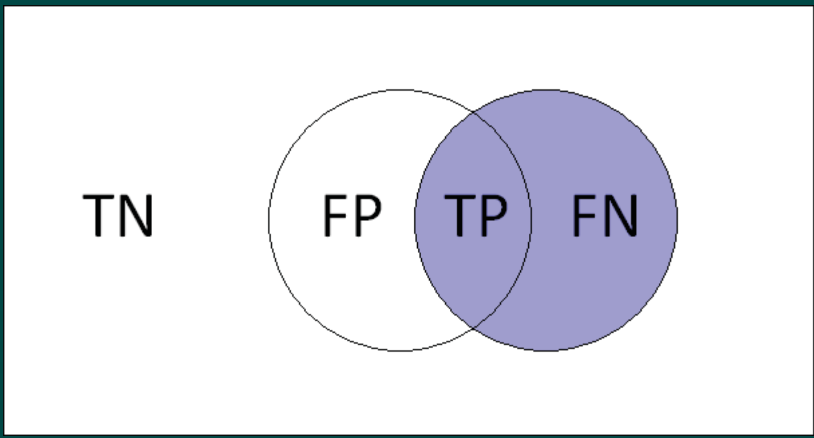
Precision, Recall, and F Measures



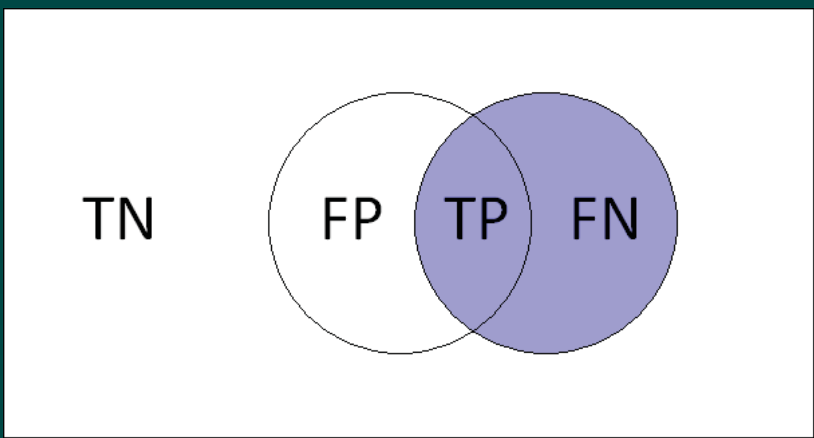
Precision, Recall, and F Measures



Precision, Recall, and F Measures



Precision, Recall, and F Measures



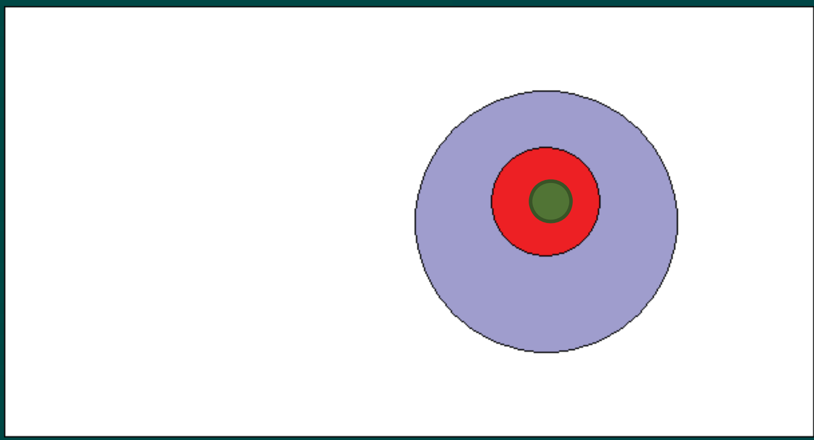
$$Precision = \frac{TP}{FP + TP} \qquad Recall = \frac{TP}{FN + TP}$$

時事問題

- COVID-19 相關議題
  - Excel 釋例

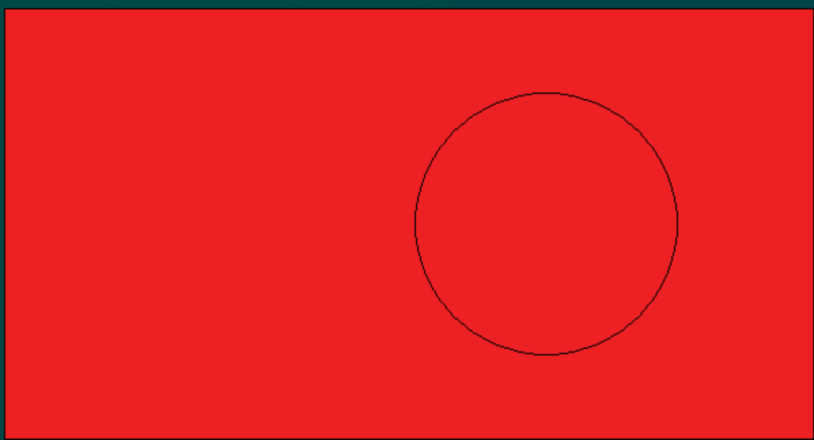
Why Two Measures

Precision vs. Recall



$$Precision = \frac{TP}{FP + TP} \qquad Recall = \frac{TP}{FN + TP}$$

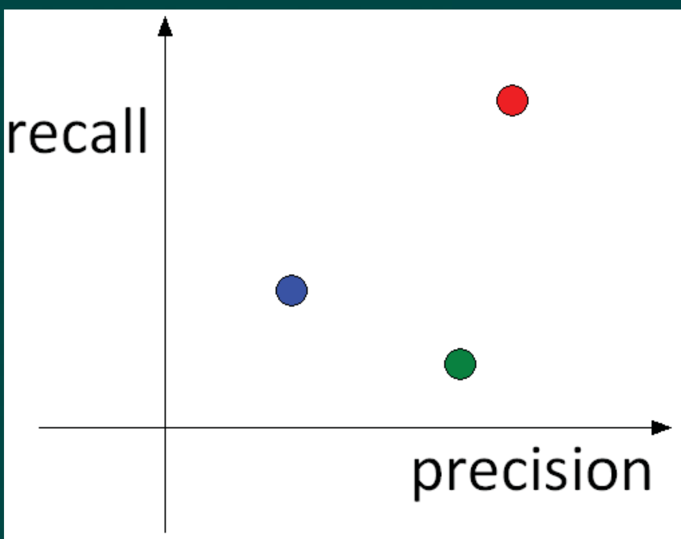
Precision vs. Recall



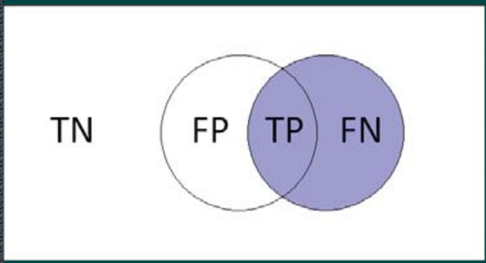
$$Precision = \frac{TP}{FP + TP} \qquad Recall = \frac{TP}{FN + TP}$$

Need One Measure for Total Ordering

Precision vs. Recall



Precision, Recall, and F Measures



$$Precision = \frac{TP}{FP + TP}$$

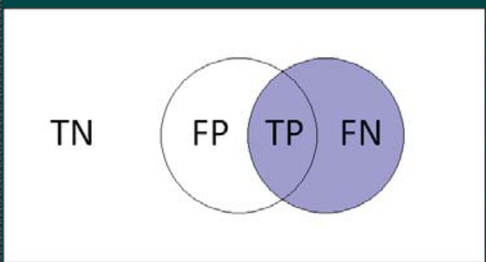
$$Recall = \frac{TP}{FN + TP}$$

$$F = \frac{Precision + Recall}{2} \ ?$$

$$F = \sqrt{Precision \times Recall} \ ?$$

$$\frac{1}{F} = \frac{1}{2}\left(\frac{1}{Precision} + \frac{1}{Recall}\right) ?$$

Precision, Recall, and F Measures



$$Precision = \frac{TP}{FP + TP}$$

$$Recall = \frac{TP}{FN + TP}$$

$$\frac{1}{F_1} = \frac{1}{2}\left(\frac{1}{Precision} + \frac{1}{Recall}\right)$$

$$\frac{1}{F_\alpha} = \alpha \frac{1}{Precision} + (1 - \alpha) \frac{1}{Recall}$$

$$\frac{1}{F_\beta} = \frac{1}{1 + \beta^2} \frac{1}{Precision} + \frac{\beta^2}{1 + \beta^2} \frac{1}{Recall}$$

MULTIPLE CLASSES

Confusion Matrix

- Table 5.4

預測類別

實際答案

	C1	C2	C3
C1	10	20	30
C2	30	20	10
C3	20	10	30

Micro and Macro Averages

- Micro precision: $\frac{10+20+30}{60+50+70}$
- Macro precision: 1/3(10/60+20/50+30/70)

預測類別

實際答案

	C1	C2	C3	Recall
C1	10	20	30	10/60
C2	30	20	10	20/60
C3	20	10	30	30/60
Precision	10/60	20/50	30/70	

More Possible Measures

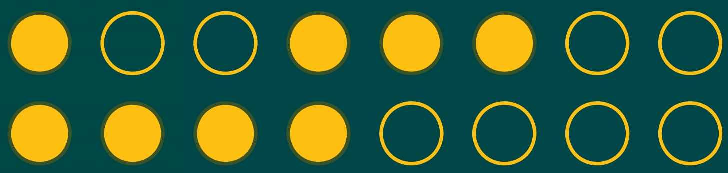
不同領域的不同選擇

- Sec. 5.8 Table 5.7 (DM4 p. 191)
- https://en.wikipedia.org/wiki/Sensitivity_and_specificity
- Sensitivity and specificity
  - https://uberpython.wordpress.com/2012/01/01/precision-recall-sensitivity-and-specificity/
- 馬偕醫院
  - http://www.mmh.org.tw/taitam/medical_edu/www/default.asp?contentID=644
  - Sensitivity (敏感度)：為有病者診斷結果為陽性的比率=真陽性率=真陽性 / 生病 = a / a+c
    - 當高靈敏診斷試驗的結果為陰性，此為未罹患此疾病相當可靠的指標
  - Specificity (特異度)：為沒病者診斷結果為陰性的比率=真陰性率=真陰性 / 健康 = d / b+d
    - 在專一性高的診斷試驗，結果陽性即表有病，因為罕見偽陽性

MAP@K

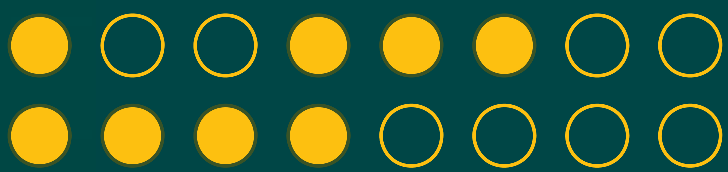
Evaluation

- Precision, recall, and F_1
- Mean Average Precision



MAP@8

- $\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{2}{4}+\frac{3}{5}+\frac{4}{6}+\frac{4}{7}+\frac{4}{8}$
- $\frac{1}{1}\times 1+\frac{1}{2}\times 0+\frac{1}{3}\times 0+\frac{2}{4}\times 1+\frac{3}{5}\times 1+\frac{4}{6}\times 1+\frac{4}{7}\times 0+\frac{4}{8}\times 0$
- $\frac{1}{4}\left(\frac{1}{1}\times 1+\frac{2}{4}\times 1+\frac{3}{5}\times 1+\frac{4}{6}\times 1\right)$
- $\frac{1}{4}\left(\frac{1}{1}\times 1+\frac{2}{2}\times 1+\frac{3}{3}\times 1+\frac{4}{4}\times 1\right)$
- $\frac{1}{2}\left(\frac{1}{4}\left(\frac{1}{1}\times 1++\frac{2}{4}\times 1+\frac{3}{5}\times 1+\frac{4}{6}\times 1\right)+\frac{1}{4}\left(\frac{1}{1}\times 1++\frac{2}{2}\times 1+\frac{3}{3}\times 1+\frac{4}{4}\times 1\right)\right)$
- $\frac{1}{2}\left(\frac{1}{4}\left(\frac{1}{1}+\frac{2}{4}+\frac{3}{5}+\frac{4}{6}\right)+\frac{1}{4}\left(\frac{1}{1}+\frac{2}{2}+\frac{3}{3}+\frac{4}{4}\right)\right)$

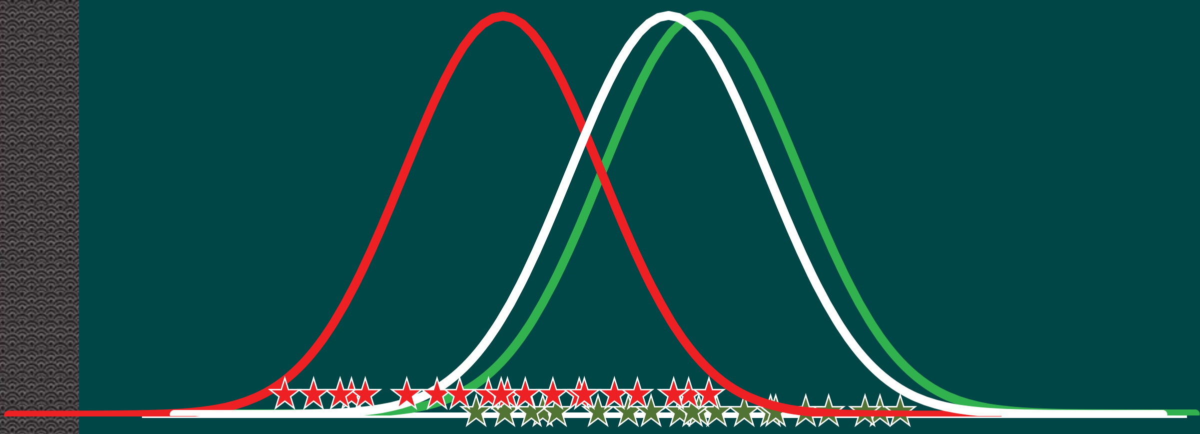


A Simple Statistical Inference

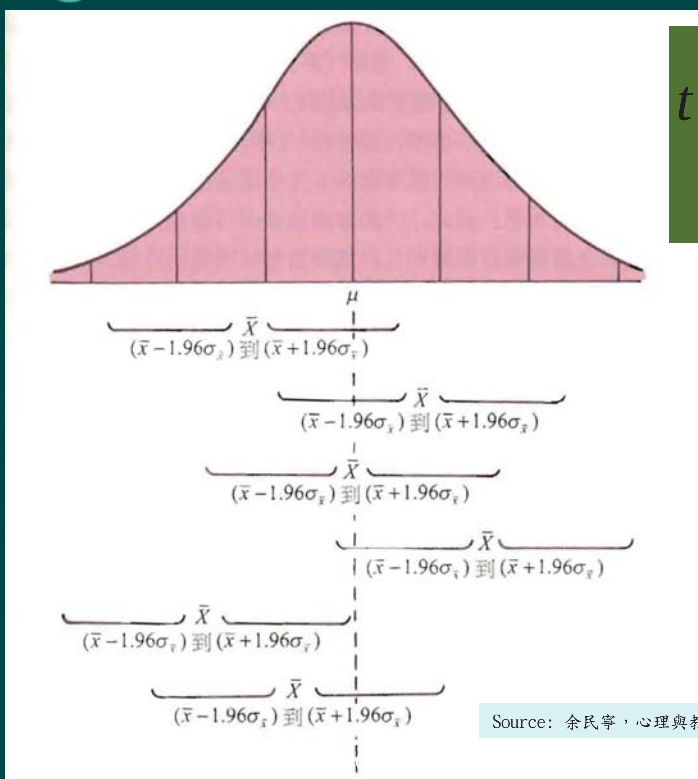
品質的比較

- DM4 Section 5.2
- Learning rules 的時候遇到的問題
  - 4/12
  - 1/1, 3/3
- 正確率
  - 75/100
  - 750/1000

Sampling



Estimating the Mean



$$t = \frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$$

Source：余民寧，心理與教育統計學，三民書局，2007。

Estimating the Mean (Cont'd)

- Central Limit Theorem

If $\overline{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ , then the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{\mathrm{n}}}$$

as $n \to \infty$, is the standard normal distribution N(0,1).

- Confidence interval for μ; σ known

Estimating the Mean (Cont'd)

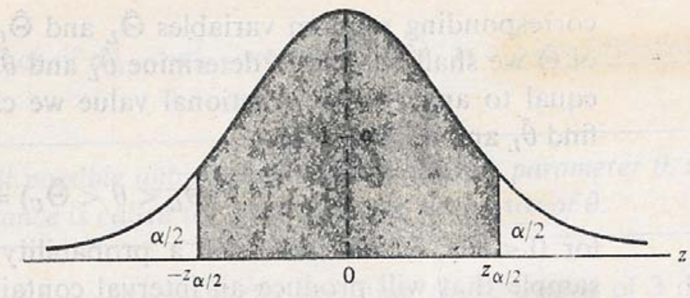


Figure 7.2 $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$

$$\Pr\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

It is quite common to set α to 0.05 or 0.01 in educational and psychological research. $Z_{\alpha/2}$ will be 1.96 and 2.81, respectively.

$$\Pr\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Details

$$\Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \qquad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

$$-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu \qquad \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu$$

$$\Pr\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

A Detailed Workout

$$\Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

Bernoulli trial:
true success rate p,
$\mu=p$ and $\sigma^2=p(1-p)$

f	n	α	$z_{\alpha/2}$	p
0.75	100	0.2	1.282	[0.691,0.801]
0.75	1000	0.2	1.282	[0.732,0.767]
0.75	100	0.1	1.645	[0.673,0.814]
0.75	1000	0.1	1.645	[0.727,0.772]

$$\Pr\left(-z_{\alpha/2} < \frac{f-p}{\sqrt{p(1-p)}/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\frac{f-p}{\sqrt{p(1-p)/n}} = z_{\alpha/2} \qquad \frac{0.75-p}{\sqrt{p(1-p)/1000}} = 1.645 \qquad p \in [0.727, 0.772]$$

Cross Validation

- Representative data
- 如何切割訓練資料和測試資料？

COST-SENSITIVE LEARNING

Various Subjects

- DM4 Section 5.8
- 機率式的預測
  - Table 5.6 (DM4 p. 184)
- Lift Chart
  - Figure 5.1 (DM4 p. 185)
- Receiver operating characteristic
  - ROC curve
  - Figure 5.3 (DM4 p. 188)

PROBABILISTIC CLASSIFICATION

QUADRATIC LOSS FUNCTION

Predicting Probabilities

- DM4 Sec. 5.7 page 177

test1 0.1 0.2 0.3 0.4 ans:1
(0.1-1)^2+0.2^2+0.3^2+0.4^2
test2 0.2 0.1 0.4 0.3 ans:3
(0.2-0)^2+(0.1)^2+(0.4-1)^2+0.3^2

$\sum_j (p_j - a_j)^2$ for a particular test instance

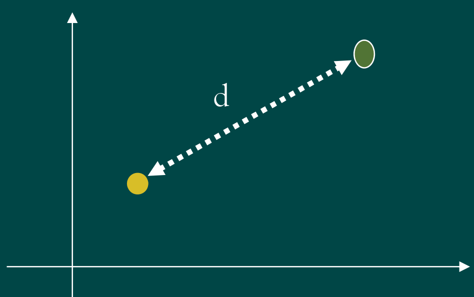
$$
\begin{aligned}
&E\left[\sum_j (p_j - a_j)^2\right] \text{ over all test instances} \\
&= \sum_j \left(E\left[p_j^2\right] - 2E\left[p_j a_j\right] + E\left[a_j^2\right]\right) \\
&= \sum_j \left(p_j^2 - 2p_j E\left[a_j\right] + E\left[a_j^2\right]\right) \\
&= \sum_j \left(p_j^2 - 2p_j p_j^* + p_j^*\right) \\
&= \sum_j \left(\left(p_j - p_j^*\right)^2 + p_j^*\left(1 - p_j^*\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
E[a_j] &= p_1^* a_j + p_2^* a_j + \cdots + p_j^* a_j + \cdots + p_n^* a_j \\
&= p_1^* 0 + p_2^* 0 + \cdots + p_j^* 1 + \cdots + p_n^* 0 \\
&= p_j^*
\end{aligned}
$$

Cross Entropy

Predicting Probabilities

- Information loss function
  - Deviation between two probabilities
  - Cross entropy (Kullback-Leibler divergence)
  - DM4 Sec. 5.7 Page 178
- Discussion



p: 0.0 0.9 0.1
q: 0.1 0.8 0.1
r: 0.2 0.7 0.1
$d(p,q)$=?

KL Divergence

- KL divergence measures the "difference" between two probability distributions, p and q.

$$D(p\|q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$
$$= \sum_{i=1}^{n} p_i (\log p_i - \log q_i)$$
$$= \sum_{i=1}^{n} (p_i \log p_i - p_i \log q_i)$$
$$= \sum_{i=1}^{n} p_i \log p_i - \sum_{i=1}^{n} p_i \log q_i$$

Therefore, the best q will minimize this quantity.

Let p be the "true" probability distribution. This quantity would be a constant for any q.

Numeric Prediction

多種選擇

- DM4 p. 195
  - Table 5.8