C26: Deep Learning

\[ \text{RNN} \rightarrow \text{forward/backprop} \]
\[ \text{CNN} \rightarrow \text{forward (1D, 2D, 3D, pooling)} \]
\[ \text{Regularization (Dropout/L2)} \]
\[ \text{forget gate (ml LSTM)} \]

C18: Representative clustering

\[ k\text{-means} \]
\[ k\text{-medoids} \rightarrow k\text{-means} \]
\[ EM \]

C15: Density-based

\[ \text{KDE density estimation} \]
\[ \text{Mean Shift} \]
\[ \text{DENCLUE} \]

C16: Spectral/graph clustering

\[ \text{ratio/normalized cut} \]
\[ \text{Laplacian, normalized Laplacian} \]
\[ \text{Spectral clustering} \]

\[ \text{MCL: Markov Clustering} \]
\[ \text{transition matrix} \]
\[ \text{random walk (1 hop walk)} \]
\[ \text{Stochastic} \]
\[ \text{Extrema Clustering} \]
C17: Clustering validation

- F-score
- NMI
- Silhouette coefficient (W: proximity/distance matrix)

Note: W in evaluation

C22: Classification Assessment

- F-score
- TPR/FPR/AUC
- Cross-validation
- T-test
- Confidence intervals

Performance metric

\[
\mathcal{D} = \{ \vec{x}_i, y_i \}
\]

\( y_i \in \{ c_1, c_2, \ldots, c_K \} \) categorical

\( \hat{y}_i = f(\vec{x}_i) \) prediction

\( d(\hat{y}_i, y_i) \) loss function

Training

Validation

Testing
\[
\text{Error rate: } \frac{1}{n} \sum_{i=1}^{n} I(\hat{y}_i \neq y_i) \\
= \frac{1}{n} \sum_{i=1}^{n} I (\hat{y}_i + y_i) \\
\text{Indicator function} \\
I(z) = \begin{cases} 
1 & \text{if } z \text{ is true} \\
0 & \text{if } z \text{ is false}
\end{cases}
\]

Accuracy = 1 - Error rate

**F-score**: Any of the harmonic means precision and recall over all the classes

\[
p_i = \text{precision for class } C_i = \text{accuracy for } C_i \\
\text{Recall for } C_i = \text{the fraction of } C_i \text{ that was predicted correctly} \\
\frac{C_i \text{ that is correctly predicted}}{C_i \text{ that is actually there}}
\]

\[
f_i = \left( \frac{1}{p_i} + \frac{1}{r_i} \right)^{-1} = \frac{2p_i r_i}{p_i + r_i}
\]

\[
F\text{-score} = \frac{1}{k} \sum_{i} f_i
\]
Confusion Matrix

\[ D_{\text{test}} \]

\[ D_i = \{ \tilde{x}_i, y_i = c_i \} \]

\[ D_2 = \{ \tilde{x}_i, y_i = c_2 \} \]

\[ \vdots \]

\[ D_k = \{ \tilde{x}_i, y_i = c_k \} \]

\[ f(x_i) = y_c \]

Predict positive

\[ R_i = \{ \tilde{x}_i, \hat{y}_i = c_i \} \]

\[ \vdots \]

\[ R_k = \{ \tilde{x}_i, \hat{y}_i = c_k \} \]

\[ \text{True} \ (y_i) \]

Predict \( \hat{y}_i \)

\[ R_{ij} = R_i \cap D_j \]

\[ R_{ik} = \{ \tilde{x} \mid y = c_j \text{ and } \hat{y} = c_i \} \]

\[ |D_i| = n_i = \text{size of class } i \in D \]

\[ |R_i| = n_i = \text{size of predicted class } (C_i) \]
\[ p_i = \frac{n_{i+}}{n_i} \quad r_i = \frac{n_{i+}}{\overline{n}_i} \]

**Precision**

**Recall**

<table>
<thead>
<tr>
<th>Predicted</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Negative</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

- **TP:** True positives
- **TN:** True negatives
- **FP:** False positives
- **FN:** False negatives

Cells associated with error

\{X_i, y_i\} \quad y_i \in \{P, N\}

\[ \overline{TP} : \quad y_i = P, \hat{y}_i = P \]
\[ \overline{TN} : \quad y_i = N, \hat{y}_i = N \]
\[ \overline{FP} : \quad \hat{y}_i = P, y_i = N \]
\[ \overline{FN} : \quad \hat{y}_i = N, y_i = P \]

\[ FN \leftarrow \text{very costly (infection detection)} \]
\[ FP \leftarrow \text{analyse him} \]
ROC curve

\[ f(x) = \hat{g}_c \]

\[ \text{take predictions from } \hat{g} \]

\[ \text{sort them in decreasing order} \]

\[ \text{Confidence} \]

\[ f_1, f_2, \ldots \]

AUC: area under curve

Ideal classifier

Higher AUC is better

True across all thresholds

Predicted

\[
\begin{array}{c|c|c}
\text{P} & \text{T} & \text{F} \\
\hline
\text{T} & \text{TP} & \text{FP} \\
\text{N} & \text{FN} & \text{TN} \\
\hline
\text{P} & \text{TP} & \text{FP} \\
\text{N} & \text{FN} & \text{TN} \\
\end{array}
\]

\[ \text{TPR: True pos rate (true for } \text{P} \) \]

\[ \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{NP}} \]

\[ \text{FPR: False pos rate } (1 - \text{true for } \text{N}) \]

\[ \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{FP}}{\text{NN}} \]
Logistic regression

$x_i \rightarrow \text{Prob for class } P, \text{true class } y_i$

$\begin{array}{c|c|c}
0.1 & P & P \\
0.8 & P & N \\
0.55 & P & N \\
0.3 & P & P \\
0.9 & P & P \\
0.2 & P & P \\
\end{array}$

sort descending order

$\begin{array}{ccccccc}
x_5 & x_2 & x_3 & x_1 & x_4 & x_6 & x_1 \\
0.9 & 0.8 & 0.55 & 0.3 & 0.2 & 0.1 & \\
\end{array}$

$\Theta = 0.5 \leftarrow \text{threshold for } P \text{ vs } N$

$\begin{array}{c|c|c|c}
2 & 1 & 0 & 2 \\
1 & 3 & 1 & 2 \\
\end{array}$

$\text{TPR} = \frac{2}{3}, \text{FPR} = \frac{2}{3}$
D: cross-validation (CV)
k-fold CV

5-fold CV:

1) Randomly shuffle the points
2) Divide data into 5 folds (partition)

D → D_1, D_2, D_3, D_4, D_5

D_1: test
D - D_1: training

D_2: test
D - D_2: 

D_5: test
$M : \text{ model}$
$
\Theta : \text{ performance matrix}$

\[
\begin{bmatrix}
D \rightarrow S \text{-folds} \\
\text{for } i = 1 \ldots S \\
D_i \text{ for testing} \\
D - D_i \text{ for training} \\
\Theta_i = \mu(D_i)
\end{bmatrix}
\]

$\Theta_1, \Theta_2, \ldots, \Theta_S$

\[
\begin{align*}
M & = \text{ mean performance} \\
\sigma_\Theta & = \text{ variance of performance}
\end{align*}
\]

$5\text{-fold} \rightarrow \text{ 5 times each run}$

$10\text{-fold} \rightarrow 10\times \text{ testing}$

$k\text{-fold} \rightarrow \text{ L00CV : leave one out CV}$

Q2: $M_1, \Theta_1 \text{ or } M_2, \Theta_2 : \text{ which is better?}$

$k\text{-fold CV}$

$1 \ 2 \ 3 \ \ldots \ 5 \ \cdots$

$M_1 : \Theta_{11} \ \Theta_{12} \ \Theta_{13} \ \Theta_{14} \ \Theta_{15} \ \rightarrow \mu_{m_1}, \sigma_{m_1}^2$

$M_2 : \Theta_{21} \ \Theta_{22} \ \Theta_{23} \ \Theta_{24} \ \Theta_{25} \ \rightarrow \mu_{m_2}, \sigma_{m_2}^2$

$0.91$
Is the diff. significant?

\[ \rightarrow \text{Is it possible just by random chance?} \]

\[ \begin{aligned}
\text{Yes} \quad &\rightarrow \text{not significant} \\
\text{No} \quad &\rightarrow \text{significant}
\end{aligned} \]

\[ \text{Pair t-test} \]

\[ M_1 : \theta_1, \theta_{12}, \theta_3, \theta_4, \theta_5 \]

\[ M_2 : \theta_1, \theta_{22}, \theta_3, \theta_4, \theta_5 \]

\[ \text{diff } \delta : \delta_1, \delta_{12}, \delta_3, \delta_4, \delta_5 \]

\[ \delta_1, \delta_{12}, \delta_3, \delta_4, \delta_5 \]

\[ \bar{\delta}_c = \frac{1}{5} \sum \delta_i = \frac{\delta_1 + \delta_{12} + \delta_3 + \delta_4 + \delta_5}{5} \]

\[ \frac{\delta_i}{\text{mean diff.}} \]

\[ \text{and var.} \]

\[ \frac{\overline{\delta_c}^2}{5} = \text{variance} \]

\[ \text{Ho: null hypothesis} \]

\[ M_0 = 0 \]

\[ \text{true mean should be} \]

\[ \text{alternative hypothesis} \]

\[ \mu_3 \neq 0 \]

\[ z\text{-score} = \frac{\bar{\delta}_c - \mu_3}{\frac{\overline{\delta_c}^2}{\sqrt{5}}} \]
Under $H_0$ $Z$-score $= \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ since $\mu_0 = 0$

look at $t$-distribution with $k-1$ degrees of freedom

$t$-distribution (like normal but more even in tail region)

$P(Z-score) \leq 0.1$ then we reject

there's a significant diff

Otherwise there is no statistically significant diff between the models