TIM 245 Lecture 7 (4/24/17)

Agenda

1) General process for supervised learning problems

2) Evaluation of classification and prediction models

3) Linear regression and introduction to prediction models

4) Work on project (Time permitting)
1. General Process for Supervised learning problems

Objective: Create a function or model that can map an input set of attributes to a target output value.

\[(X_1, X_2, \ldots, X_m) \rightarrow \Phi \rightarrow \hat{Y}_i\]

Attribute Values
Model

The model, \( \Phi \), is created through a supervised learning process:

- Dataset
- Training Dataset
- Test dataset
- Learning Algorithm
- Model \( \Phi \)
- Evaluation

\[
(X_1, X_2, \ldots, X_m) \rightarrow \hat{Y}_i
\]

\[
(Y_1, Y_2, \ldots, Y_n)
\]
Two general applications of supervised learning

**Descriptive Modeling**
Use \( \hat{D} \) to understand relationships in the historical data
(Focus is interpretability)

**Predictive Modeling**
Use \( \hat{D} \) to predict \( y_i \) for a new instance \((x_{i1}, x_{i2}, \ldots, x_{in})\)
(Focus is accuracy)

Key issue in Descriptive Modeling: Correlation vs Causation

Four possible cases for \( \hat{D} : X \rightarrow Y \)

1) Causal Link

\[
\begin{align*}
X & \rightarrow Y \\
Y & \rightarrow X
\end{align*}
\]
No sure which one

2) Hidden Cause

\[
\begin{align*}
\hat{D} & \rightarrow Y \\
X & \rightarrow Y \\
Z & \rightarrow X
\end{align*}
\]
Z is not in the data set
3) Confounding Factor

Z → Y
X

4) Coincidence (Noise)

X Y

Typically addressed through:
- Careful with attribute selection
- Consulting with subject matter experts
- Randomized experiments (A/B testing)

Key issue in Predictive Modeling: overfitting vs underfitting

underfitting

overfitting

just right
Typically addressed through:
- Model regularization (Lasso, Ridge)
- Careful evaluation of the accuracy
- Cross validation
- Ensemble methods (Bagging, Boosting)
Evaluation of classification and prediction models

Classification models are evaluated using a confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Award</th>
<th>Predicted</th>
<th>No Award</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Award</td>
<td>True</td>
<td>False</td>
<td>Negative</td>
</tr>
<tr>
<td>Actual No</td>
<td>False</td>
<td>True</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Measures: Accuracy, Precision, Recall, f-measure, g-mean

Example: Accuracy = \( \frac{TP + TN}{TP + TN + FP + FN} \)

Prediction models are evaluated based on the error

\( e_i = y_i - \hat{y}_i \) (residual)

Measures: Sum of Squared Error (SSE), Mean Absolute Error (MAE), Mean Average Percent Error (MAPE)

Example: \( SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \)
(3) Linear Regression

Three general types of relationships between X and Y

Positive Correlation: Values move in same direction

\[ \uparrow X \uparrow Y \text{ or } \downarrow X \downarrow Y \]

Negative Correlation: Values move in opposite directions

\[ \uparrow X \downarrow Y \text{ or } \downarrow X \uparrow Y \]

No Correlation

\[ \uparrow X \downarrow Y \text{ or } \downarrow X \uparrow Y \text{ or } \uparrow X \downarrow Y \text{ or } \downarrow X \uparrow Y \]

Linear regression estimates the type and strength of linear relationships in the training data set, i.e., best straight line fit
Example: HWY 17 Commute times vs number of cars on the road

Minutes

150
120
90
60
30

100 200 300 400 500
Cars on the road

error or residual

General form of a linear regression model

\[ \hat{y} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi} \]

dependent variable
intercept
coefficients
independent variables

Coefficients indicate the strength and type of the relationship:

Magnitude: Strength of the relationship
- \( \approx 0 \): no correlation

Sign: Type of relationship
- \(+\): positively correlated
- \(-\): negatively correlated
Simple Linear Regression: linear regression with only a single independent variable or attribute

\[ \hat{y} = \beta_0 + \beta_j X_{ij} \quad \text{(think } y = mx + b) \]

Process for Simple linear Regression:

1) Separate the dataset into training dataset and test dataset

2) Plot data for \( j = 1 \) (A1)
3) Estimate $\beta_0$ (intercept) and $\beta_1$ (slope) on the training dataset for $j=1$ ($A_1$) using least squares

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_{i1} - \bar{A}_1)(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_{i1} - \bar{A}_1)^2}
$$

$$
\beta_0 = \bar{Y} - \beta_1 \bar{A}_1
$$

where $\bar{A}_1$ and $\bar{Y}$ are the means of $A_1$ and $Y$, respectively.

4) Calculate the sum of squared error on the test dataset

$$
SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2
$$

5) Repeat for each attribute $A_j$ in the dataset ($j=2, 3, \ldots, m$)

6) Select the model with the lowest error

Provides a good baseline for more sophisticated linear models