False Negatives and False Positives in IDSs

CSE 4471, Intro to Information Security
Adam C. Champion

Adapted from Prof. Srinivasan Parthasarathy’s data mining slides
Confusion Matrix (1)

• Suppose IDS gets \( n \) packets: \( k \) attack, \( n - k \) benign (where \( 0 \leq k \leq n \))
• IDS thinks: \( l \) attack, \( n - l \) benign (where \( 0 \leq l \leq n \))
• Confusion matrix helps us understand how IDS performed (correct/incorrect classification)
• Context: positive means malicious, negative means benign
Confusion Matrix (2)

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class = Positive</td>
<td>Class = Positive</td>
<td>(TruePos)</td>
<td>True positive</td>
</tr>
<tr>
<td>Class = Negative</td>
<td>Class = Negative</td>
<td>(FalseNeg)</td>
<td>False negative</td>
</tr>
<tr>
<td>Class = Positive</td>
<td>(FalsePos)</td>
<td>False positive</td>
<td></td>
</tr>
<tr>
<td>Class = Negative</td>
<td>(TrueNeg)</td>
<td>True negative</td>
<td></td>
</tr>
</tbody>
</table>

- **True positive**: IDS classified MALICIOUS packet as *malicious*
- **True negative**: IDS classified BENIGN packet as *benign*
- **False negative**: IDS classified MALICIOUS packet as *benign*
- **False positive**: IDS classified BENIGN packet as *malicious*
Confusion Matrix (3)

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class = Positive</td>
<td>Class = Positive</td>
<td>(TruePos) True positive</td>
</tr>
<tr>
<td>Class = Negative</td>
<td>Class = Negative</td>
<td>(FalseNeg) False negative</td>
</tr>
</tbody>
</table>

- False negative rate \( \text{FalseNegRate} = \frac{\text{FalseNeg}}{\text{FalseNeg} + \text{TruePos}} \)
- False positive rate \( \text{FalsePosRate} = \frac{\text{FalsePos}}{\text{FalsePos} + \text{TrueNeg}} \)

\( k \), \( n - k \), \( l \), \( n - l \)
Example 1

- \( n = 1000 \) packets
- \( k = 250 \) real attacks
- \( l = 360 \) attacks (according to the IDS)
- IDS detects 0 \textit{real} attacks

- What’s the false positive rate?
- What’s the false negative rate?
Example 2

- $n = 1000$ packets
- $k = 250$ real attacks
- $l = 360$ attacks (according to the IDS)
- IDS detects 250 *real* attacks

- What’s the false positive rate?
- What’s the false negative rate?
Example 3

- \( n = 1000 \) packets
- \( k = 250 \) real attacks
- \( l = 360 \) attacks (according to the IDS)
- IDS detects 125 \textit{real} attacks

- What’s the false positive rate?
- What’s the false negative rate?