Exercises in Machine Learning
Playing with Kernels

Zdeněk Žabokrtský, Ondřej Bojar
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University, Prague

Tue Mar 31, 2015
Outline

- Linear Kernel: \( k(x, y) = x \cdot y \)
- Polynomial Kernel: \( k(x, y) = (\gamma \cdot x \cdot y + \text{coeff0})^{\text{degree}} \)
- RBF Kernel: \( k(x, y) = \exp(-\gamma \|x - y\|^2); \gamma > 0 \)
  ... including their parameters
- Cross-validation Heatmap
- Multi-class SVM
  - For the PAMAP-easy dataset.
  - Regularization parameters.
  - Inseparable classes.

Regularization (C) in linear SVM

\[ k(x, y) = x \cdot y \]

(Linear kernel = no kernel)

The parameter \( C \) in (linear) SVM:
- sets the weight of the sum of slack variables.
- serves as a regularization parameter.
- controls the number of support vectors.

<table>
<thead>
<tr>
<th>( C ) for Errors</th>
<th>Number of points considered</th>
<th>Bias</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Many</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Few</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

Think \( C \) for Variance.
SVM Linear $C=0.1$
SVM Linear $C=0.2$
SVM Linear $C=0.5$
SVM Linear $C=1$
SVM Linear $C=5$
SVM Linear $C=10$
SVM Linear $C=20$
SVM Linear C=50
SVM Linear $C=100$
Polynomial Kernel

\[ k(x, y) = (\gamma \cdot x \cdot y + \text{coeff0})^{\text{degree}} \]
SVM Poly (degree 1)
SVM Poly (degree 2)
SVM Poly (degree 3)
SVM Poly (degree 4)
SVM Poly (degree 5)
SVM Poly (degree 6)
SVM Poly (degree 7)
SVM Poly (degree 8)
SVM Poly (degree 9)
SVM Poly (degree 3, gamma 0.05)
SVM Poly (degree 3, gamma 0.1)
SVM Poly (degree 3, gamma 0.2)
SVM Poly (degree 3, gamma 0.5)
SVM Poly (degree 3, gamma 0.7)
SVM Poly (degree 3, gamma 1)
SVM Poly (degree 3, gamma 2)
SVM Poly (d=3, g=0.5, coef=-2.0)
SVM Poly (d=3, g=0.5, coef=-1.0)
SVM Poly ($d=3$, $g=0.5$, $coef=-0.50$)
SVM Poly (d=3, g=0.5, coef=0)
SVM Poly (d=3, g=0.5, coef=0.5)
SVM Poly \((d=3, \ g=0.5, \ \text{coef}=1)\)
SVM Poly \( (d=3, \ g=0.5, \ \text{coef}=2) \)
RBF Kernels

\[ k(x, y) = \exp(-\gamma \| x - y \|^2); \gamma > 0 \]

- Each training point creates its bell.
- Overall shape is the sum of the bells.
- Kind of “all nearest neighbours”.
# RBF Kernel Parameters

<table>
<thead>
<tr>
<th>$C$</th>
<th>Decision Surface</th>
<th>Model</th>
<th>Bias</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Smooth</td>
<td>Simple</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Peaked</td>
<td>Complex</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>gamma</th>
<th>Affected Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>can be far from training examples</td>
</tr>
<tr>
<td>High</td>
<td>must be close to training examples</td>
</tr>
</tbody>
</table>

- Does higher gamma lead to higher variance?
- Choice critical for SVM performance.
- Advised to use GridSearchCV for $C$ and gamma:
  - exponentially spaced probes
  - wide range
SVM RBF (C=0.05, gamma=2)
SVM RBF (C=0.1, gamma=2)
SVM RBF ($C=0.2$, $\gamma=2$)
SVM RBF (C=0.5, gamma=2)
SVM RBF ($C=0.6$, $\text{gamma}=2$)
SVM RBF ($C=0.7$, $\text{gamma}=2$)
SVM RBF \((C=1, \text{ gamma}=2)\)
SVM RBF (C=2, gamma=2)
SVM RBF (C=0.5, gamma=0.05)
SVM RBF (C=0.5, gamma=0.1)
SVM RBF (C=0.5, gamma=0.2)
SVM RBF (C=0.5, gamma=0.5)
SVM RBF (C=0.5, gamma=0.7)
SVM RBF (C=0.5, gamma=1)
SVM RBF (C=0.5, gamma=2)
SVM RBF (C=0.5, gamma=5)
SVM RBF ($C=0.5$, $\text{gamma}=10$)
Cross-validation Heatmap

Multi-class SVM

Two implementations in scikit-learn:

- **SVC**: one-against-one
  - \( n(n - 1)/2 \) classifiers constructed
  - supports various kernels, incl. custom ones

- **LinearSVC**: one-vs-the-rest
  - \( n \) classifiers trained
Default View (every 200)

- SVC with linear kernel
- SVC with RBF kernel (gamma 0.7)
- SVC with polynomial (degree 3) kernel
- LinearSVC (linear kernel)

regularization: C=1.0
Default View (every 300)

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: C=1.0
Default View (every 400)

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: C=1.0
Regularization $C=0.5$

- SVC with linear kernel
- SVC with RBF kernel (gamma 0.7)
- SVC with polynomial (degree 3) kernel
- LinearSVC (linear kernel)

Regularization: $C=0.5$
Regularization $C=1$

- **SVC with linear kernel**
- **SVC with RBF kernel (gamma 0.7)**
- **SVC with polynomial (degree 3) kernel**
- **LinearSVC (linear kernel)**

regularization: $C=1.0$
Regularization $C=5$

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: $C=5.0$
Regularization $C=10$

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: $C=10.0$
Regularization $C=20$

- SVC with linear kernel
- SVC with RBF kernel (gamma 0.7)
- SVC with polynomial (degree 3) kernel
- LinearSVC (linear kernel)

regularization: $C=20.0$
Regularization $C=50$

- SVC with linear kernel
- SVC with RBF kernel (gamma 0.7)
- SVC with polynomial (degree 3) kernel
- LinearSVC (linear kernel)

Regularization: $C=50.0$
Regularization $C=500$

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: $C=500.0$
Regularization $C=5000$

$C = 5000.0$
Inseparable classes 12,13 (every 200)

SVC with linear kernel
SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel
LinearSVC (linear kernel)

regularization: C=1.0
Inseparable classes 12,13 (every 100)

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: C=1.0
Inseparable classes 12,13 (every 80)

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: C=1.0
Inseparable classes 12,13 (every 60)

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: C=1.0
Inseparable classes 12,13 (every 55)

SVC with linear kernel

SVC with RBF kernel (gamma 0.7)

SVC with polynomial (degree 3) kernel

LinearSVC (linear kernel)

regularization: C=1.0
Task and Homework

▶ Required minimum: For PAMAP-Easy:
  ▶ Cross-validate to choose between linear, poly and RBF.
  ▶ Create the heatmap for RBF.
  ▶ Use the GridSearchCV to find the best $C$ and gamma.
▶ Optional: $n$-fold cross-validation for GridSearchCV:
  ▶ Use just the training data, perhaps subsampled.
  ▶ Run GridSearchCV $n$ times, using each $1/n$ as the test set and the remaining $9/n$ as the training data.
  ▶ Report min, max, average and standard deviation of achieved $C$, gamma and test error.

▶ Try to implement it as a generic command-line utility.
▶ Just call it from the Makefile.

Due: 2 weeks from now, i.e. April 14.