Support Vector Machines and Radon’s Theorem

Henry Adams
Colorado State University
Joint with Brittany Story, Elin Farnell
AATRN, www.aatrn.net, 1-2 live talks per week
YouTube: 3,800 subscribers, 24 hours watched per day
Radon's Theorem  Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.
Radon’s Theorem. Given \( n+2 \) points in \( \mathbb{R}^n \), there exist two disjoint subsets whose convex hulls intersect.
Radon's Theorem. Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.
Radon's Theorem. Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.
Radon's Theorem. Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.
Radon’s Theorem. Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.
Radon's Theorem. Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.

$\mathbb{R}^2$:

$\mathbb{R}^1$:
**Radon’s Theorem**  Given \( n+2 \) points in \( \mathbb{R}^n \), there exist two disjoint subsets whose convex hulls intersect.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.

The classical application of Radon's theorem to SVMs is to show that the Vapnik-Chervonenkis (VC) dimension of affine separators in $\mathbb{R}^n$ is $n+1$. 
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.

**Theorem**  A separating hyperplane is optimal iff the projections of the support vectors give a labelling satisfying Radon's theorem.
Support Vector Machine (SVM)

A support vector machine finds the hyperplane separating two classes of data with the widest margin.

Theorem  A separating hyperplane is optimal if and only if the projections of the support vectors give a labelling satisfying Radon’s theorem.
With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?
With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?
With randomly chosen linearly separable data, what is the expected \# of support vector configurations of each type?

1,000 trials in $\mathbb{R}^3$

<table>
<thead>
<tr>
<th>Support vectors</th>
<th>Configuration</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td><img src="image" alt="Diagram 1/1" /></td>
<td>554</td>
</tr>
<tr>
<td>2/1</td>
<td><img src="image" alt="Diagram 2/1" /></td>
<td>367</td>
</tr>
<tr>
<td>2/2</td>
<td><img src="image" alt="Diagram 2/2" /></td>
<td>47</td>
</tr>
<tr>
<td>3/1</td>
<td><img src="image" alt="Diagram 3/1" /></td>
<td>32</td>
</tr>
</tbody>
</table>
With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?

<table>
<thead>
<tr>
<th>Distance between classes</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support vectors: 1/1</td>
<td>279</td>
<td>554</td>
<td>758</td>
</tr>
<tr>
<td>Support vectors: 2/1</td>
<td>458</td>
<td>367</td>
<td>221</td>
</tr>
<tr>
<td>Support vectors: 2/2</td>
<td>145</td>
<td>47</td>
<td>9</td>
</tr>
<tr>
<td>Support vectors: 3/1</td>
<td>118</td>
<td>32</td>
<td>12</td>
</tr>
</tbody>
</table>
With randomly chosen linearly separable data, what is the expected number of support vector configurations of each type?

<table>
<thead>
<tr>
<th>Distance between classes</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support vectors: 1/1</td>
<td>279</td>
<td>554</td>
<td>758</td>
</tr>
<tr>
<td>Support vectors: 2/1</td>
<td>458</td>
<td>367</td>
<td>221</td>
</tr>
<tr>
<td>Support vectors: 2/2</td>
<td>145</td>
<td>47</td>
<td>9</td>
</tr>
<tr>
<td>Support vectors: 3/1</td>
<td>118</td>
<td>32</td>
<td>12</td>
</tr>
</tbody>
</table>
Questions

Firey dice problem?

Kernel SVM?

Spherical or ellipsoidal SVM?

Tverberg's theorem and multiclass SVM?

Soft margin SVM?