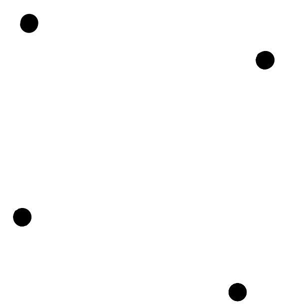
Support Vector Machines and Radon's Theorem

Henry Adams
Colorado State University
Joint with Brittany Story, Elin Farnell

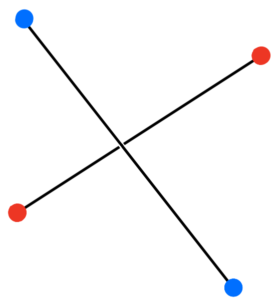
arXiv:2011.00617

AATRN, www.aatrn.net, 1-2 live talks per week
YouTube: 3,800 subscribers, 24 hours watched per day

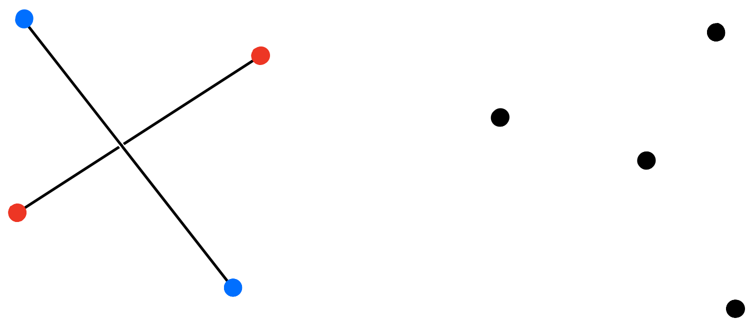
<u>Radon's Theorem</u> Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.



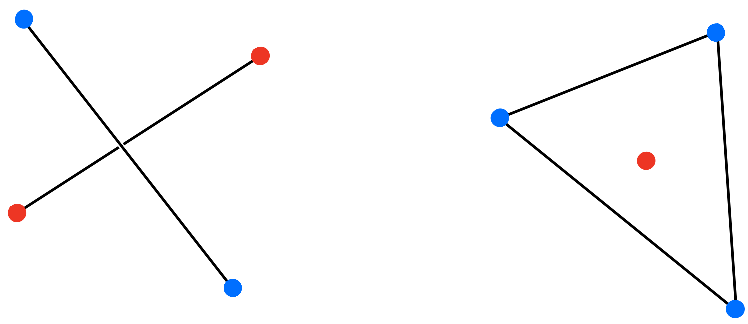
Radon's Theorem Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.



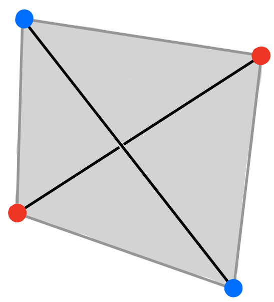
Radon's Theorem Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.

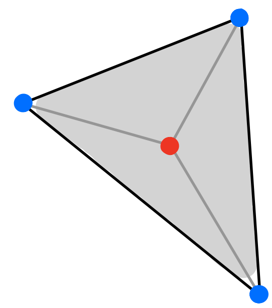


Radon's Theorem Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.



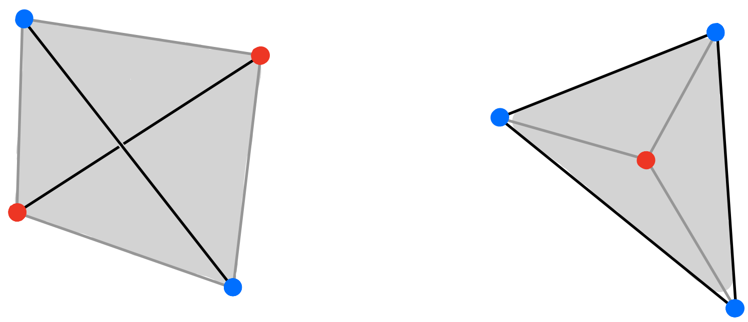
Radon's Theorem Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.



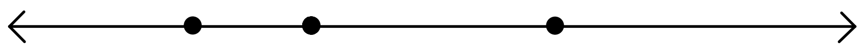


<u>Radon's Theorem</u> Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.

$\mathbb{R}^2$:

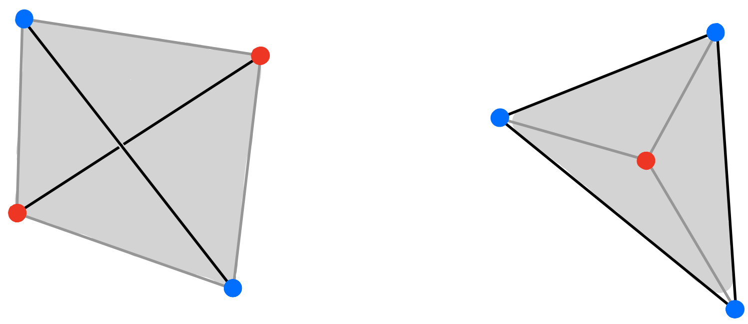


$\mathbb{R}^1$:

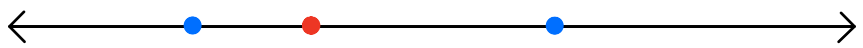


Radon's Theorem Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.

$\mathbb{R}^2$:

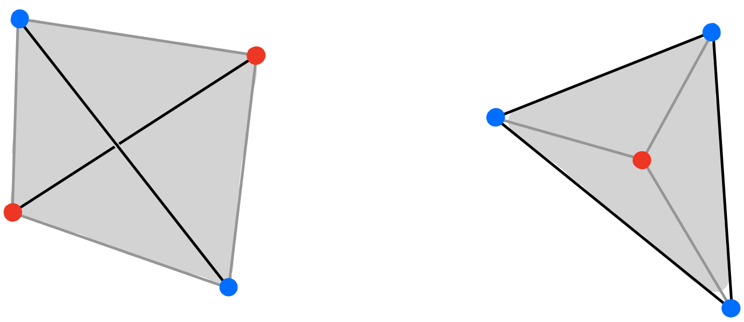


$\mathbb{R}^1$:

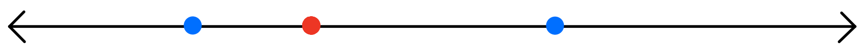


<u>Radon's Theorem</u> Given $n+2$ points in $\mathbb{R}^n$, there exist two disjoint subsets whose convex hulls intersect.

$\mathbb{R}^2$:



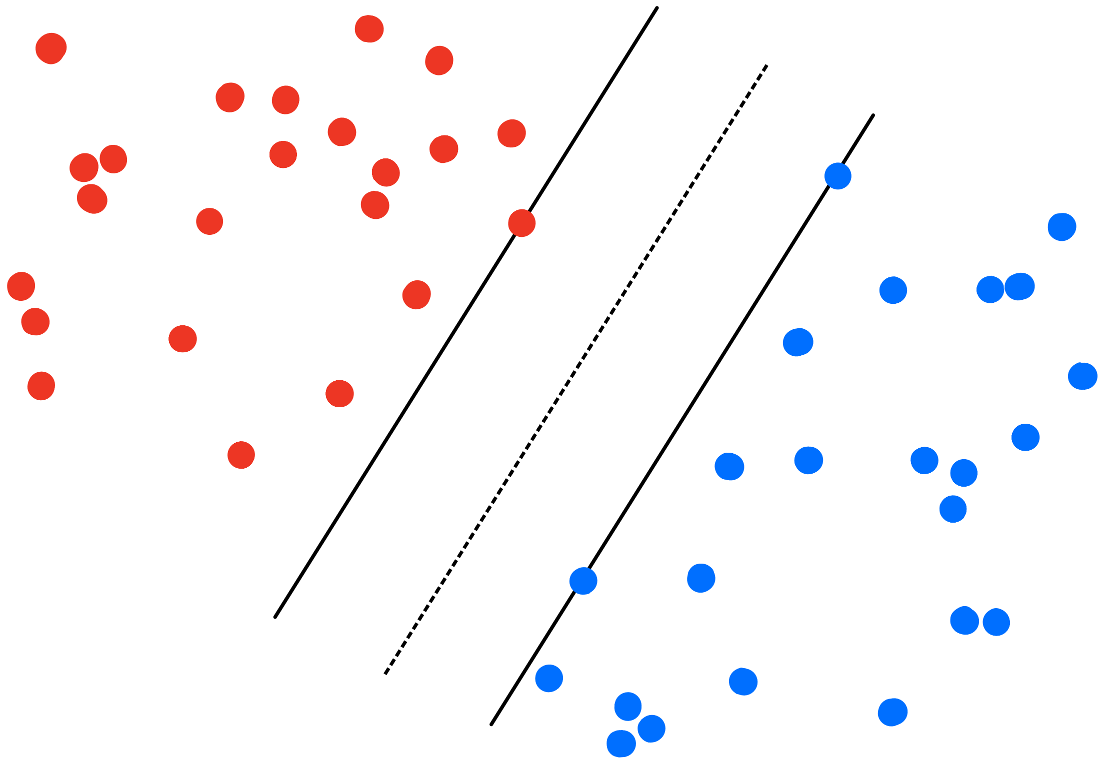
$\mathbb{R}^1$:



$\mathbb{R}^0$:

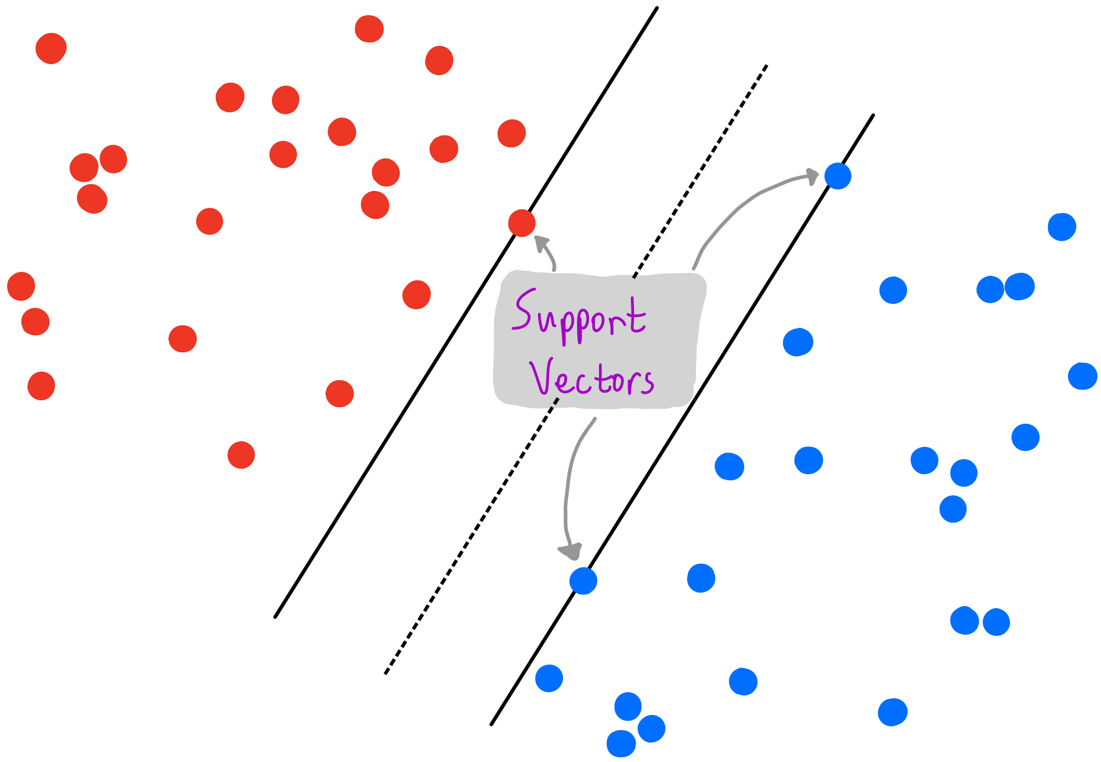
Support Vector Machine (SVM)

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



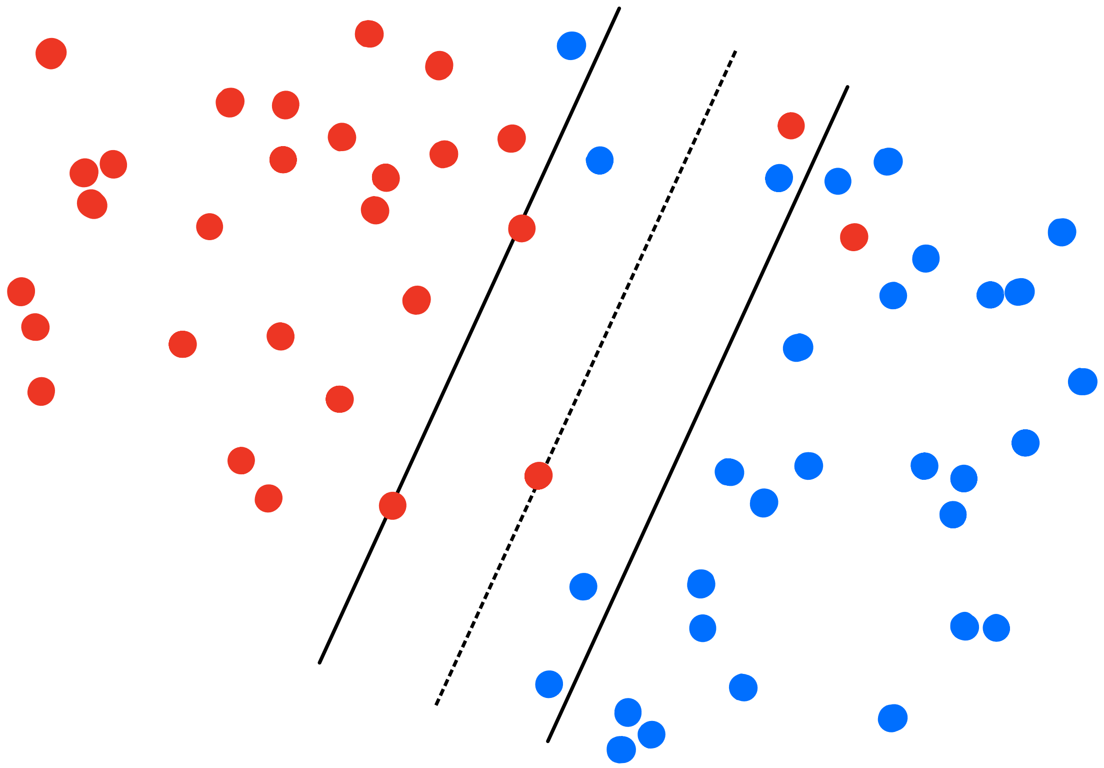
Support Vector Machine (SVM)

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



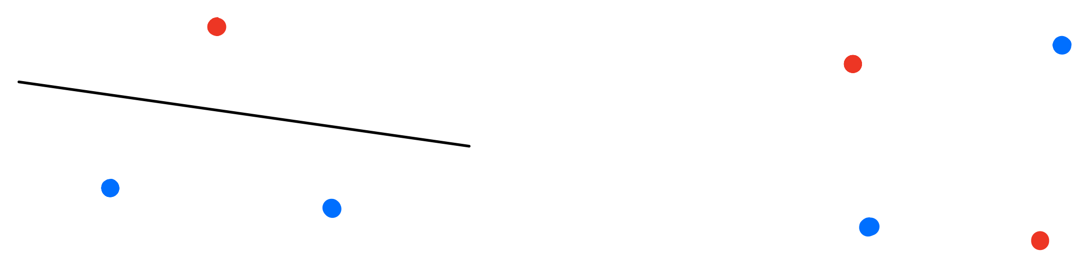
Support Vector Machine (SVM)

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



Support Vector Machine (SVM)

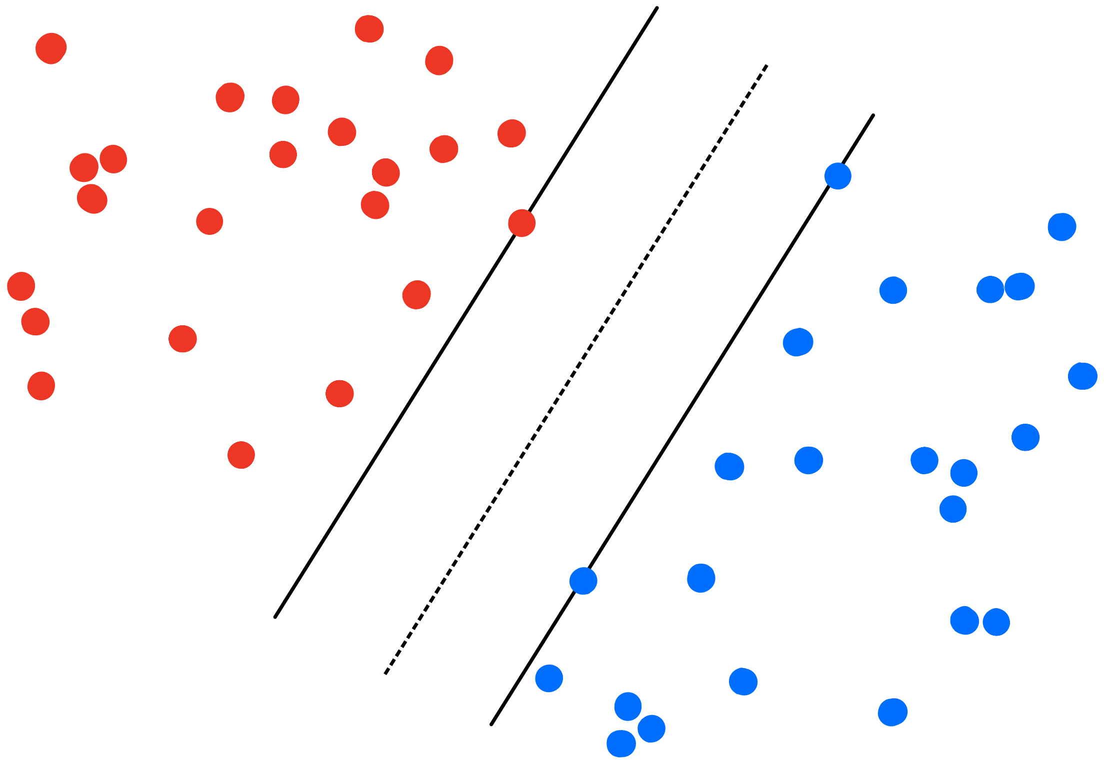
A support vector machine finds the hyperplane separating two classes of data with the widest margin.



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 $n+1$.

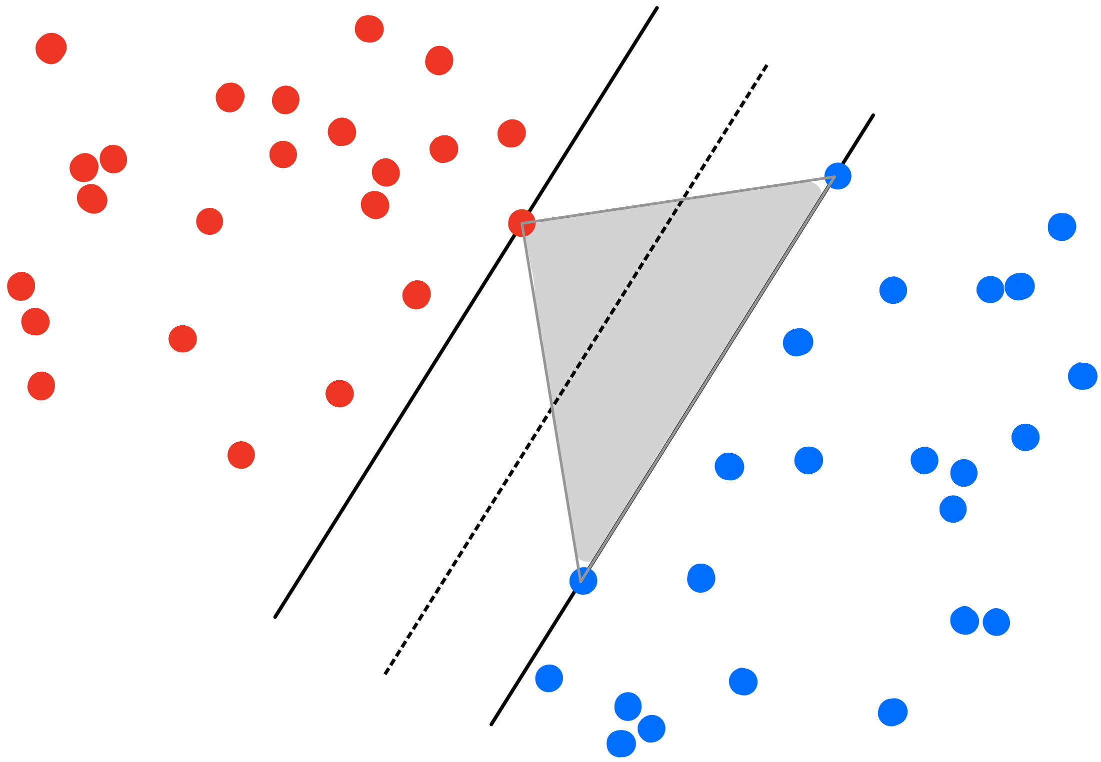
Support Vector Machine (SVM)

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



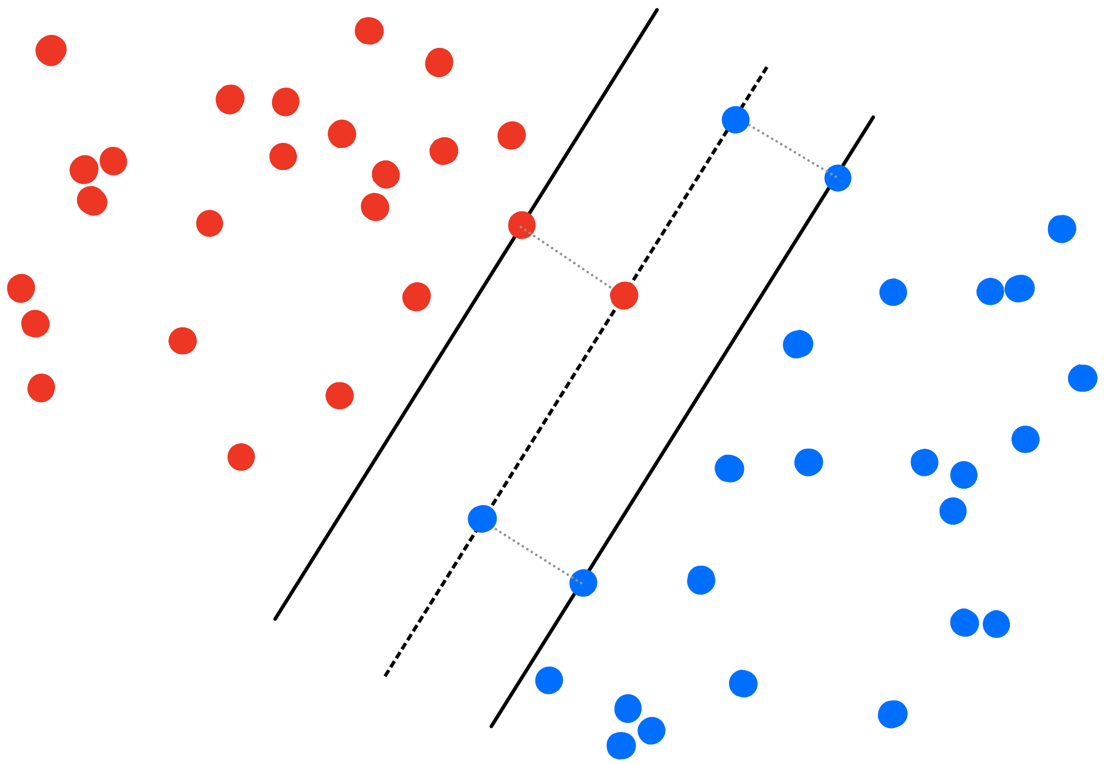
Support Vector Machine (SVM)

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



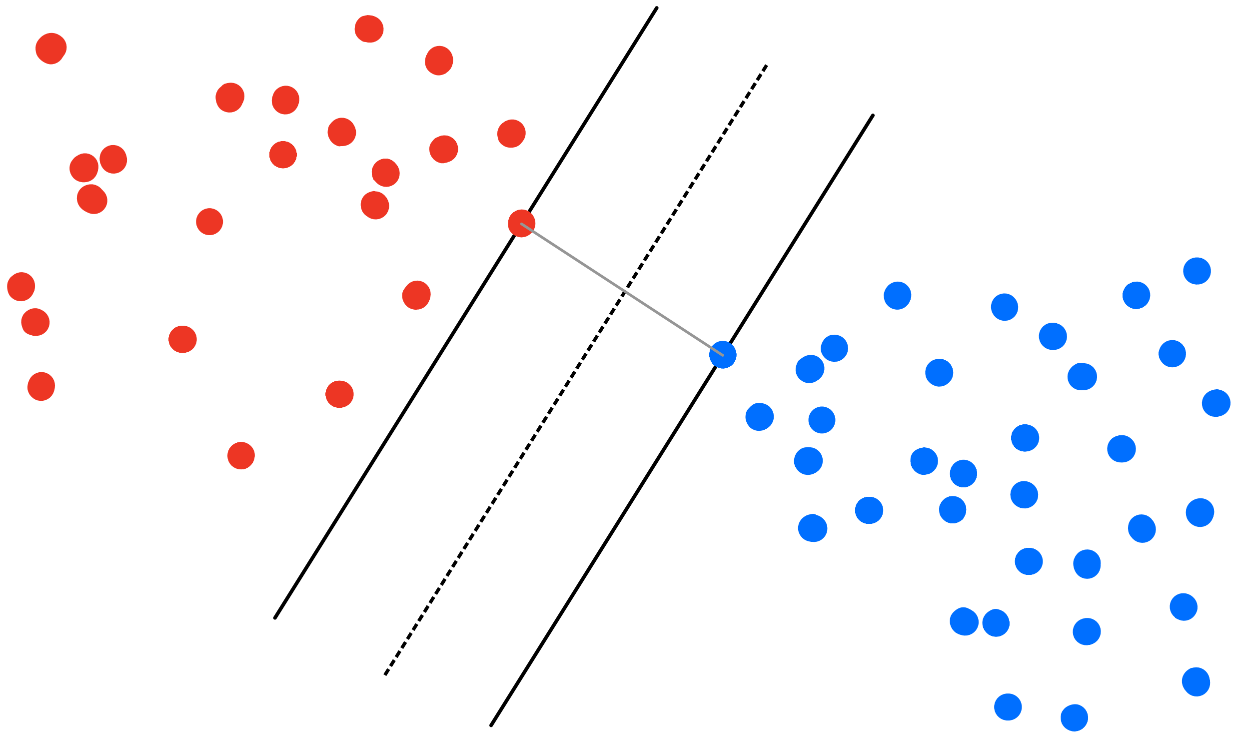
Support Vector Machine (SVM)

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



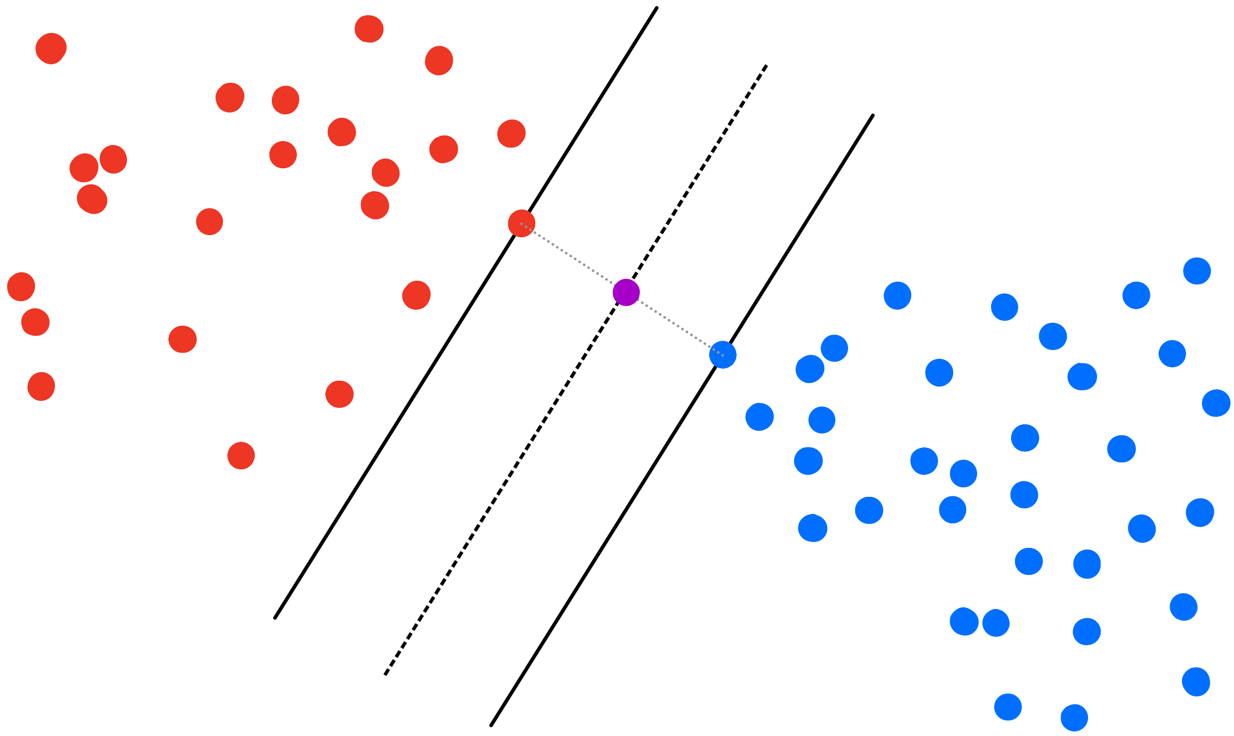
Support Vector Machine (SVM)

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



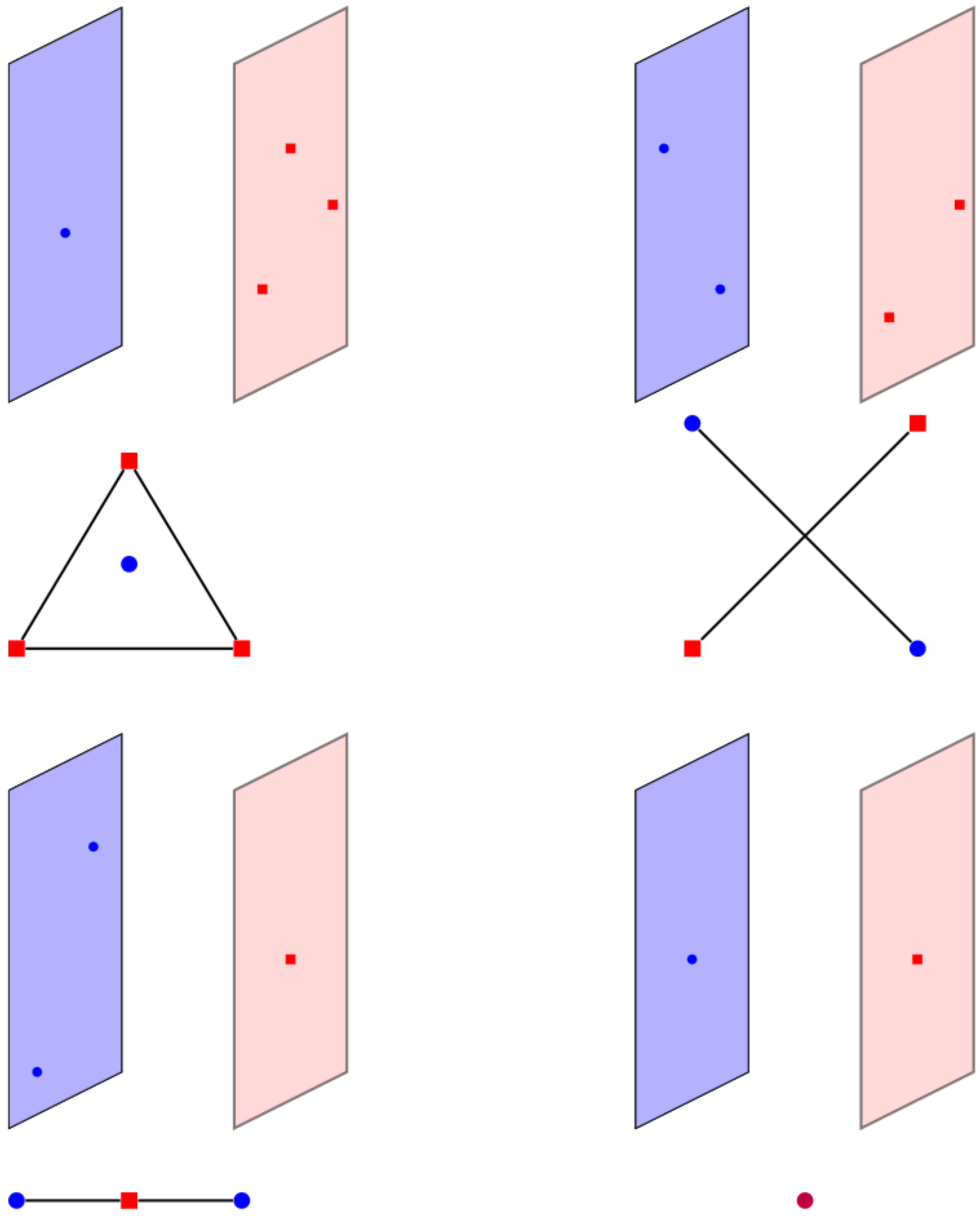
Support Vector Machine (SVM)

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



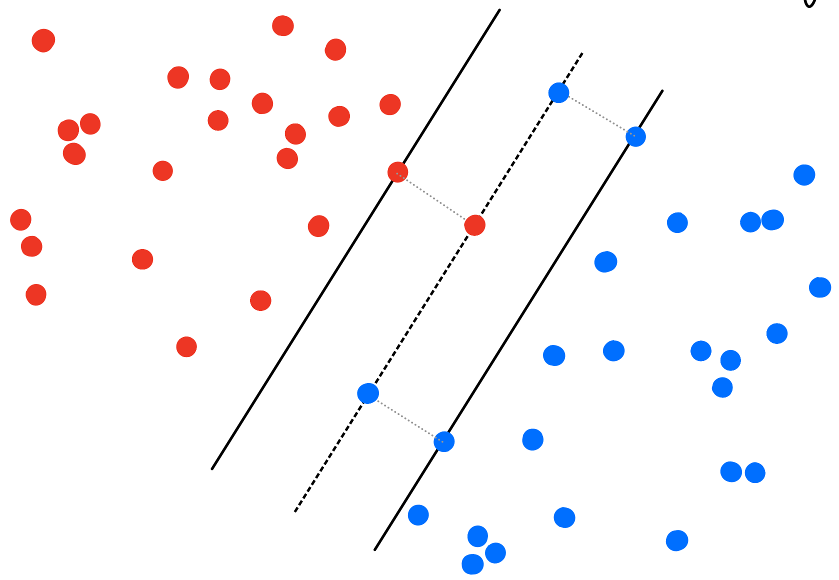
Support Vector Machine (SVM)

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



Support Vector Machine (SVM)

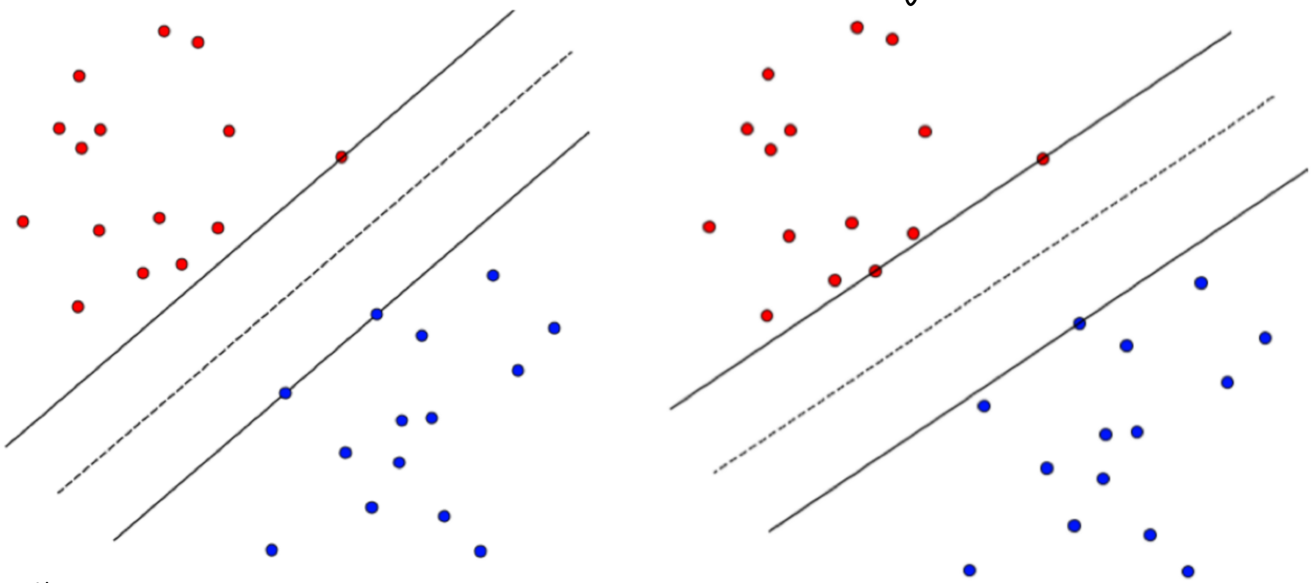
A support vector machine finds the hyperplane separating two classes of data with the widest margin.



Theorem A separating hyperplane is optimal $\Longleftrightarrow$ 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.

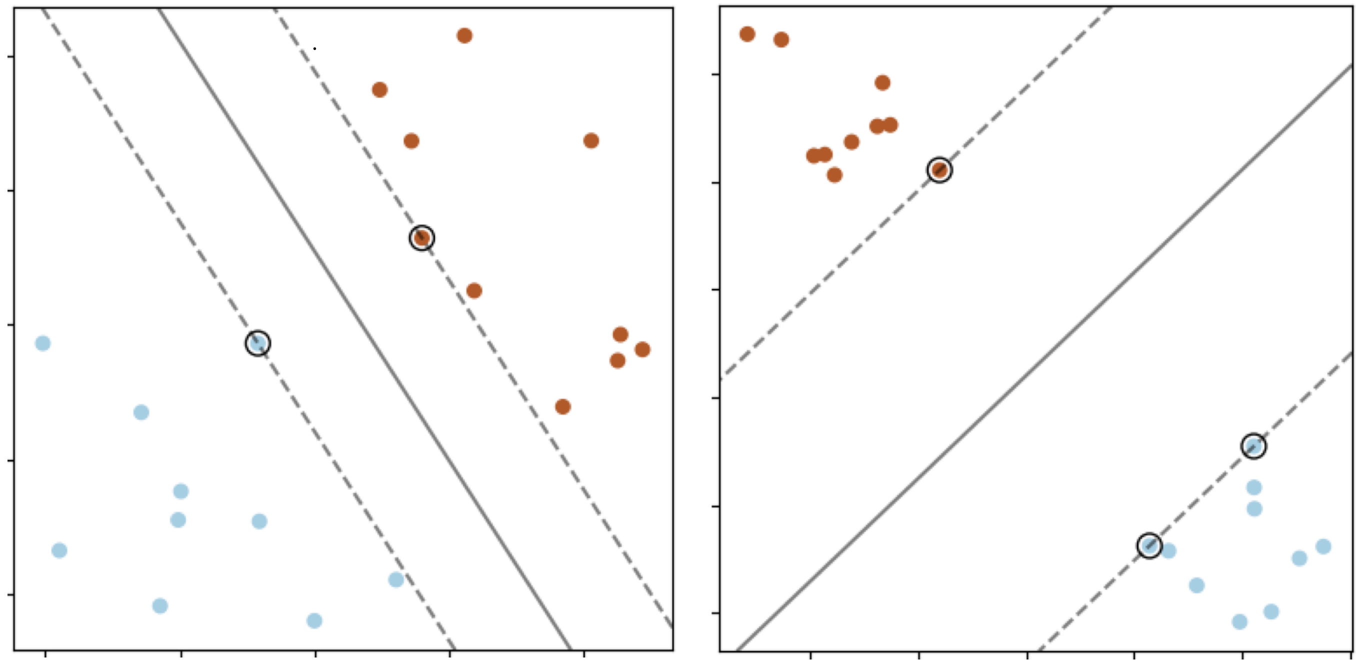


Theorem A separating hyperplane is optimal

$\Longleftrightarrow$ the projections of the support vectors give a labelling satisfying Radon's theorem.

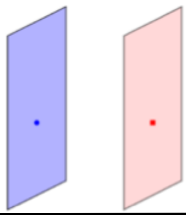
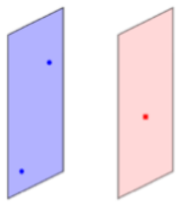
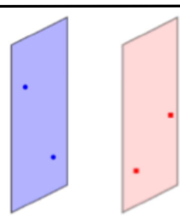
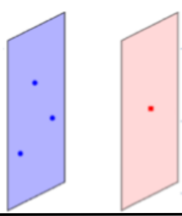
With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?

With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?



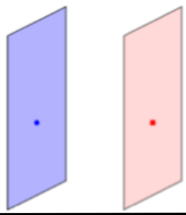
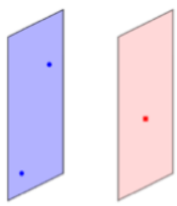
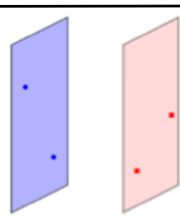
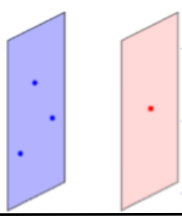
With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?

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

Support vectors: 1 / 1 	554
Support vectors: 2 / 1 	367
Support vectors: 2 / 2 	47
Support vectors: 3 / 1 	32

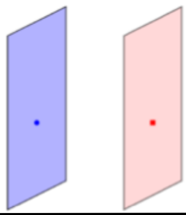
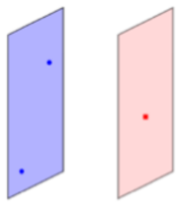
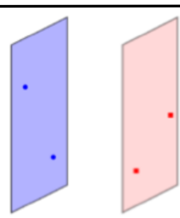
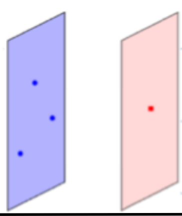
With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?

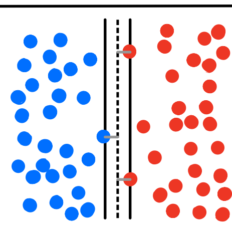
1,000 trials in $\mathbb{R}^3$

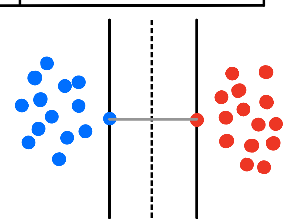
Distance between classes	Small	Medium	Large
Support vectors: 1/1 	279	554	758
Support vectors: 2/1 	458	367	221
Support vectors: 2/2 	145	47	9
Support vectors: 3/1 	118	32	12

With randomly chosen linearly separable data, what is the expected # of support vector configurations of each type?

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

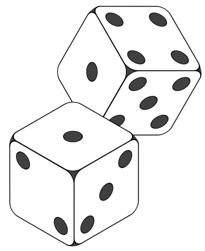
Distance between classes	Small	Medium	Large
Support vectors: 1/1 	279	554	758
Support vectors: 2/1 	458	367	221
Support vectors: 2/2 	145	47	9
Support vectors: 3/1 	118	32	12



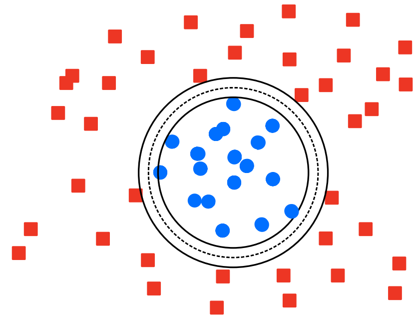


Questions

Firey dice problem?



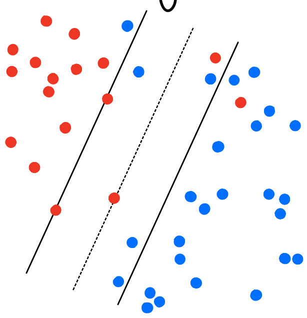
Kernel SVM?



Spherical or ellipsoidal SVM?

Tverberg's theorem and multiclass SVM?

Soft margin SVM?



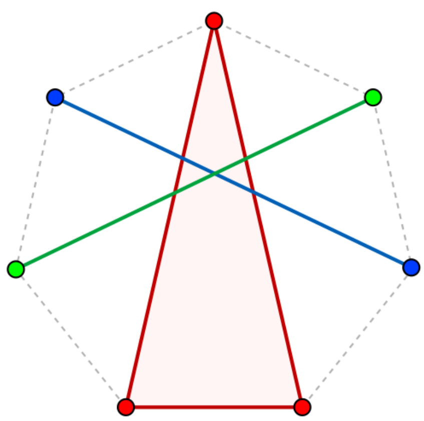


Image: Wikipedia, David Eppstein