Predictive Crime Modeling

Chicago - A Case Study

Katherine Schinkel

Scenario

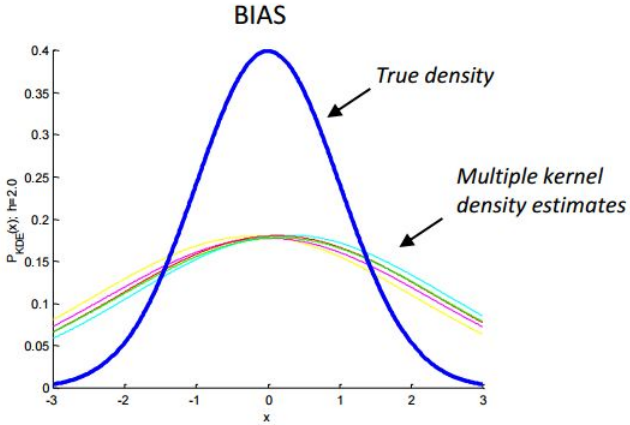
The Chicago Police Department (CPD) has limited resources (officers, patrol cars, etc.). In an effort to better deploy these resources, the CPD would like to target areas that are predicted to have high crime concentration.

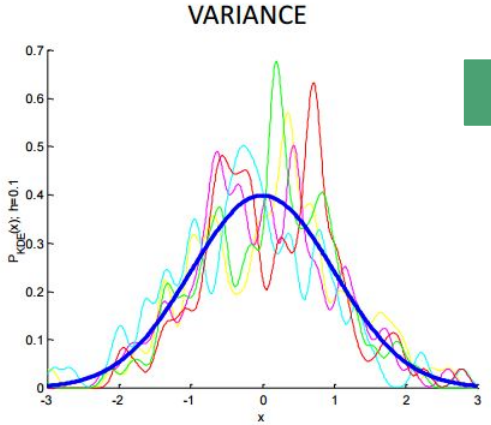
Kernel Density Estimation

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where ***K*** is the kernel function and ***h***
is the smoothing parameter (bandwidth)

Large *h*



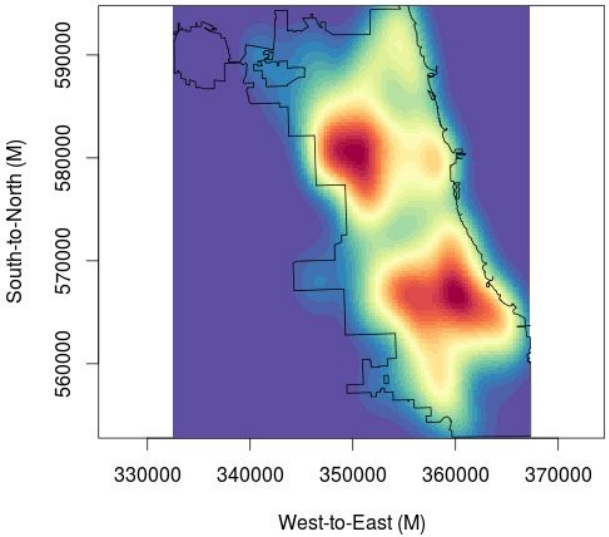


Small *h*

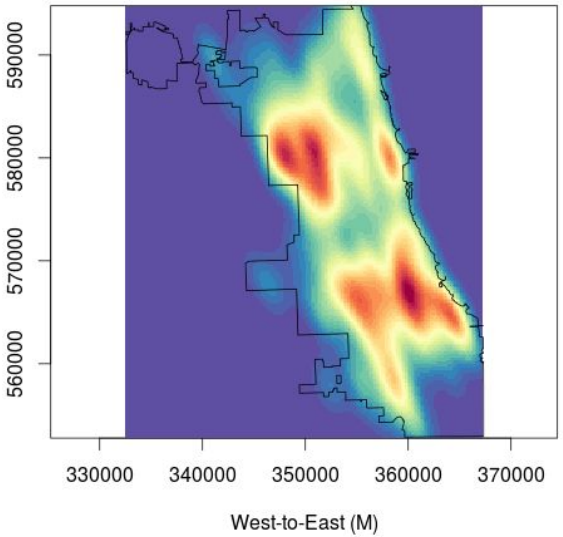
http://research.cs.tamu.edu/prism/lectures/pr/pr_l7.pdf

Kernel Density Estimation

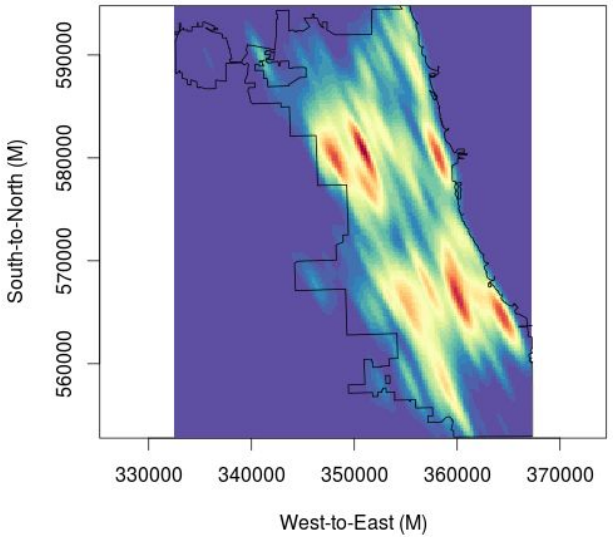
Large *h*



**Low Variance,
High Bias**



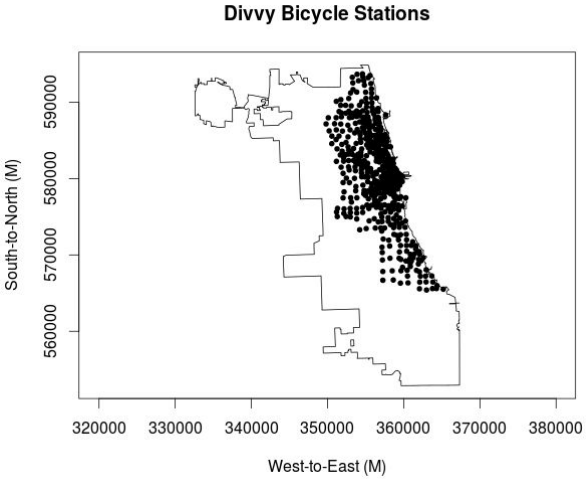
Small *h*

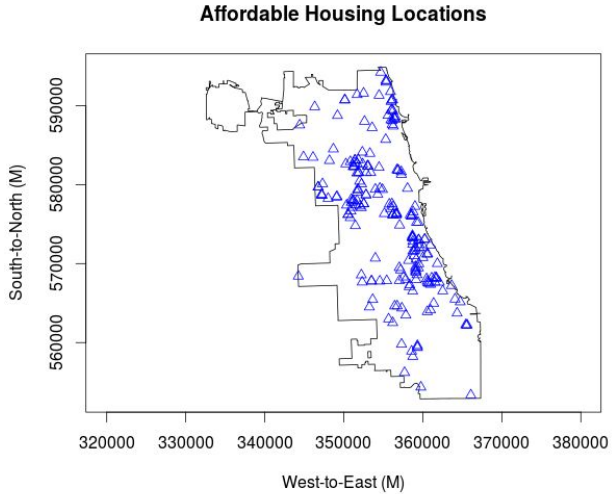


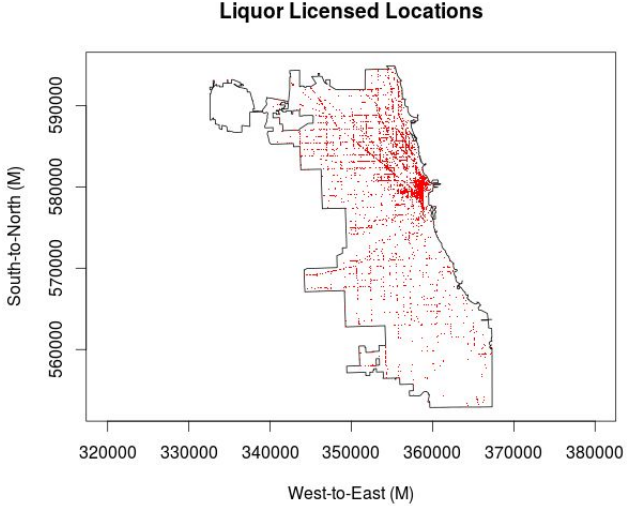
**High Variance,
Low Bias**

Spatial Parameters

Minimum distance from an assault location to...



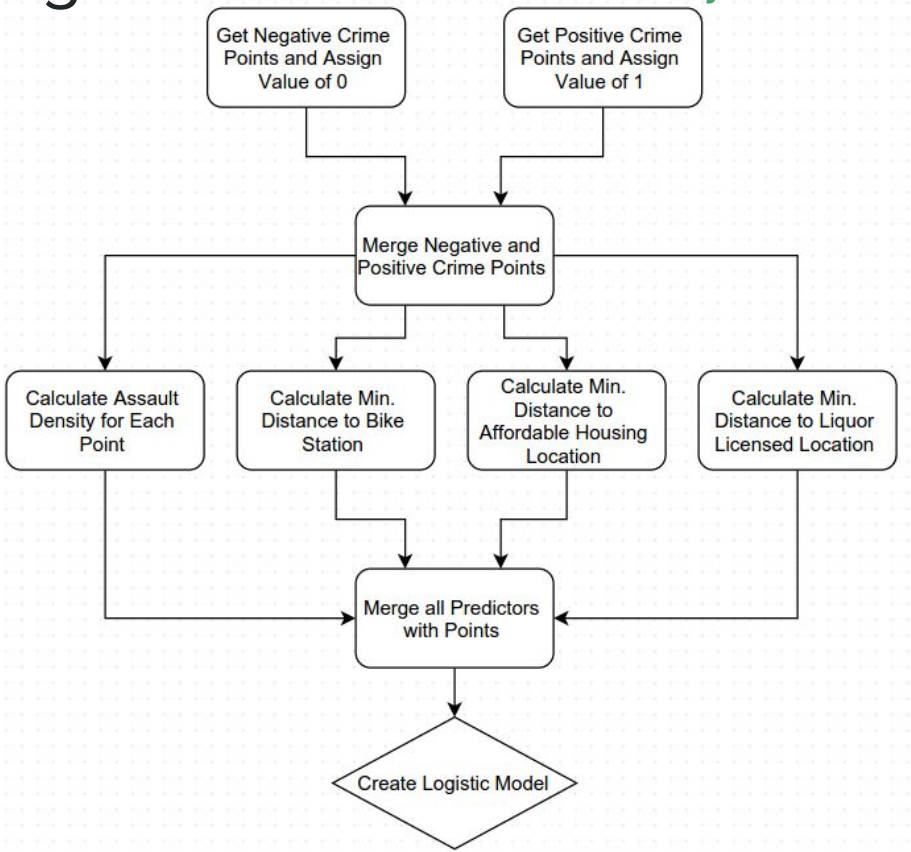




Build a Training Data Set

From February

Estimating February KDE from January Crime Points



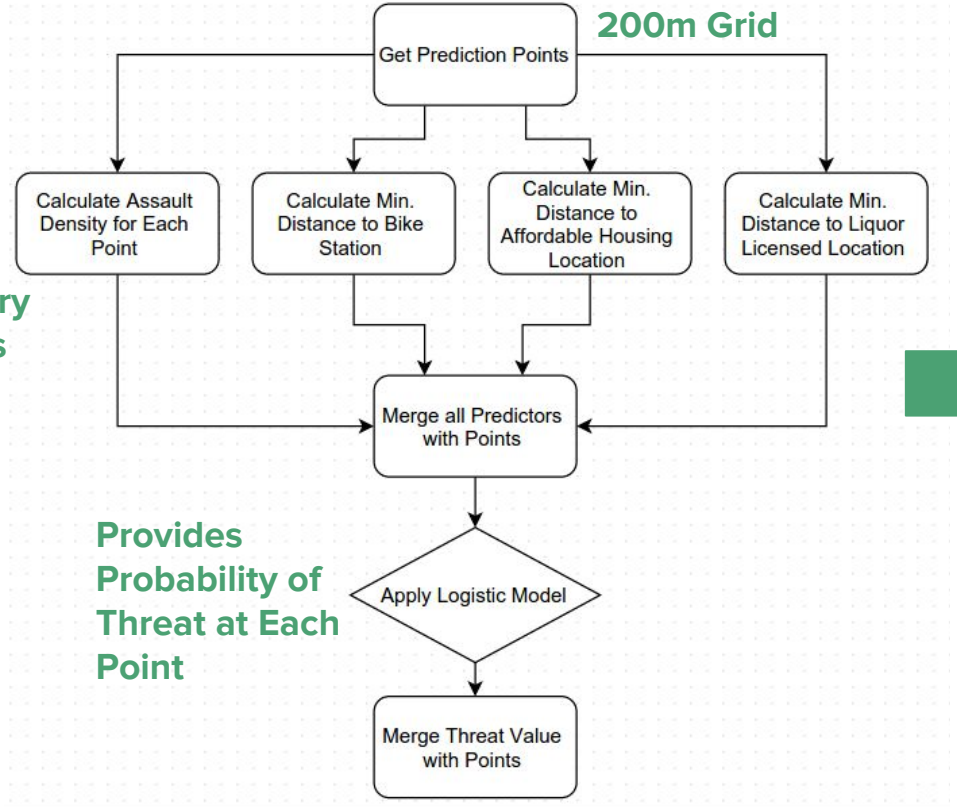
Build a Training Set - Results

	response	x	y	assault.density	bike.min.distance	liquor.min.distance	housing.min.distance
36133	0	355377.3	594275.4	1.134490e-09	804.009381	333.2899222	701.252761
36304	0	354777.3	594475.4	1.319335e-09	791.597094	38.5786120	275.442688
36305	0	354977.3	594475.4	1.280002e-09	862.556889	169.0004584	394.258182
36306	0	355177.3	594475.4	1.188713e-09	970.248842	118.6137652	561.257832
36307	0	355377.3	594475.4	1.054665e-09	999.037050	185.9284403	744.702084
36477	0	354577.3	594675.4	1.253620e-09	967.618820	227.8076056	468.106390
36478	0	354777.3	594675.4	1.255907e-09	986.744450	223.4178531	467.647765
36479	0	354977.3	594675.4	1.205782e-09	1044.530056	265.1715426	546.136679
36480	0	355177.3	594675.4	1.106858e-09	1135.086545	104.0005127	676.635878
36481	0	355377.3	594675.4	9.698520e-10	1195.708564	176.9640254	835.108946
141721	1	357230.3	554378.4	5.301562e-10	12155.353075	1274.7153375	1882.045414
14076	1	358654.9	566054.1	2.695123e-09	665.127002	302.9782443	472.124463
141961	1	358320.9	561293.1	2.667389e-09	5166.704588	445.6389323	1813.260435

Build a Prediction Set

Equally Spaced Points from 200m Grid

Estimating Prediction Points KDE from February Crime Points



Provides Probability of Threat at Each Point

Results

	x	y	threat
311	359777.3	553075.4	0.034714596
312	359977.3	553075.4	0.033074257
313	360177.3	553075.4	0.029662405
314	360377.3	553075.4	0.025413353
315	360577.3	553075.4	0.020999304
316	360777.3	553075.4	0.016960993
317	360977.3	553075.4	0.017740310
318	361177.3	553075.4	0.017601444
319	361377.3	553075.4	0.016551133

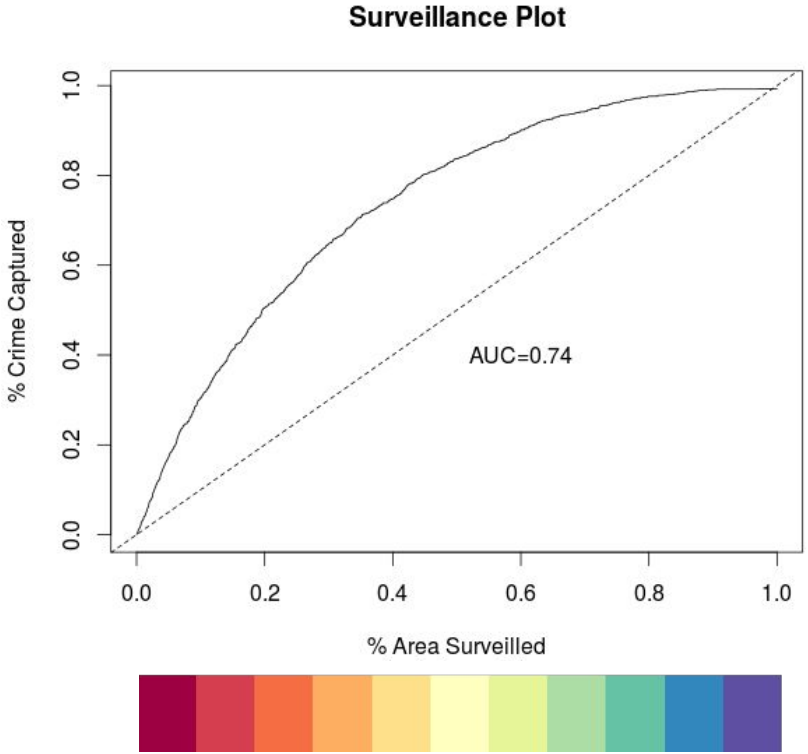
How do we evaluate the Logistic Model?

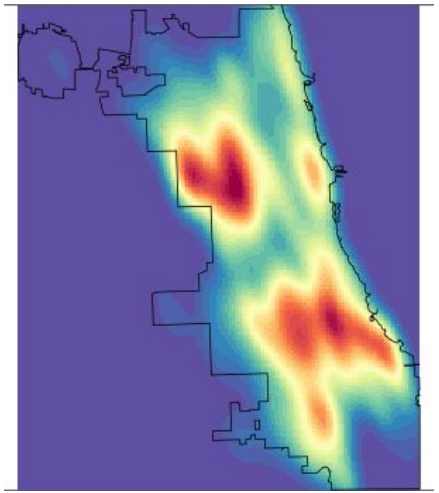
Threat Predictions

Prediction Data Set

Evaluation Points

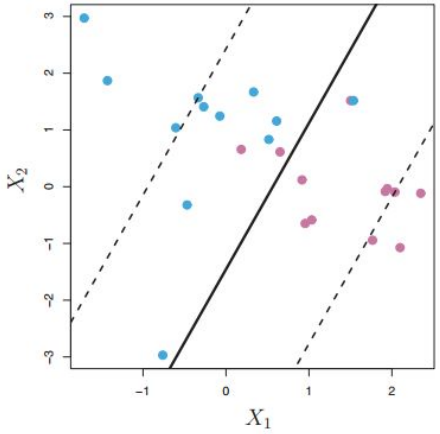
March Crime Points



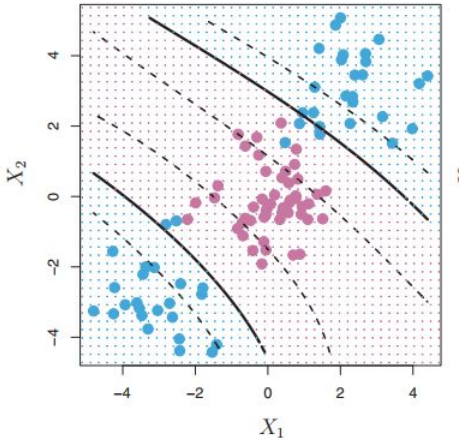


What if the relationships are not linear?

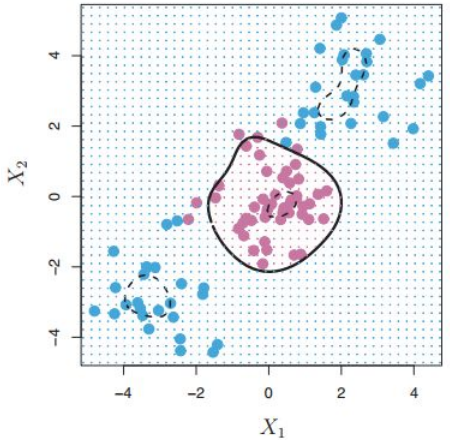
Linear SVM



Polynomial SVM



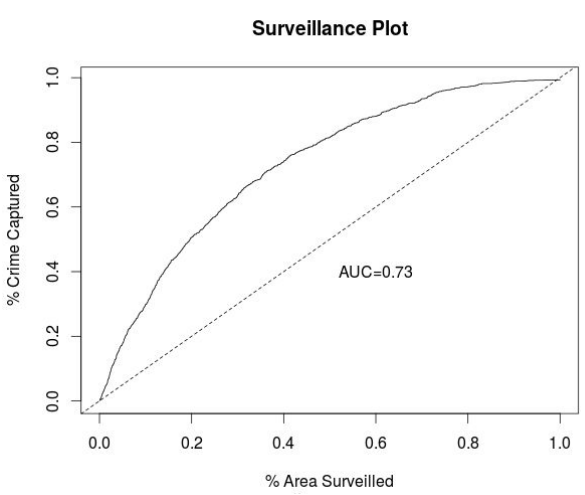
Radial Basis Function Kernel SVM



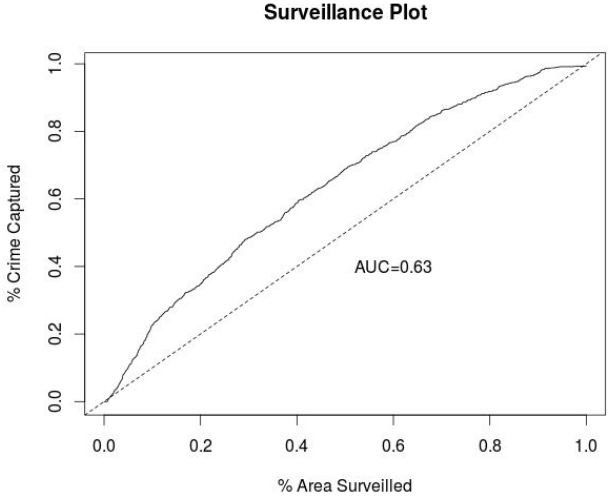
An Introduction to Statistical Learning. James, Witten, Hastie, and Tibshirani.

Support Vector Machine Modeling

Linear SVM



Quadratic SVM



Radial Basis Function Kernel SVM