Geospatial Data Science
Content Block II: *Techniques*
Lecture 7
Models for predicting spatial patterns

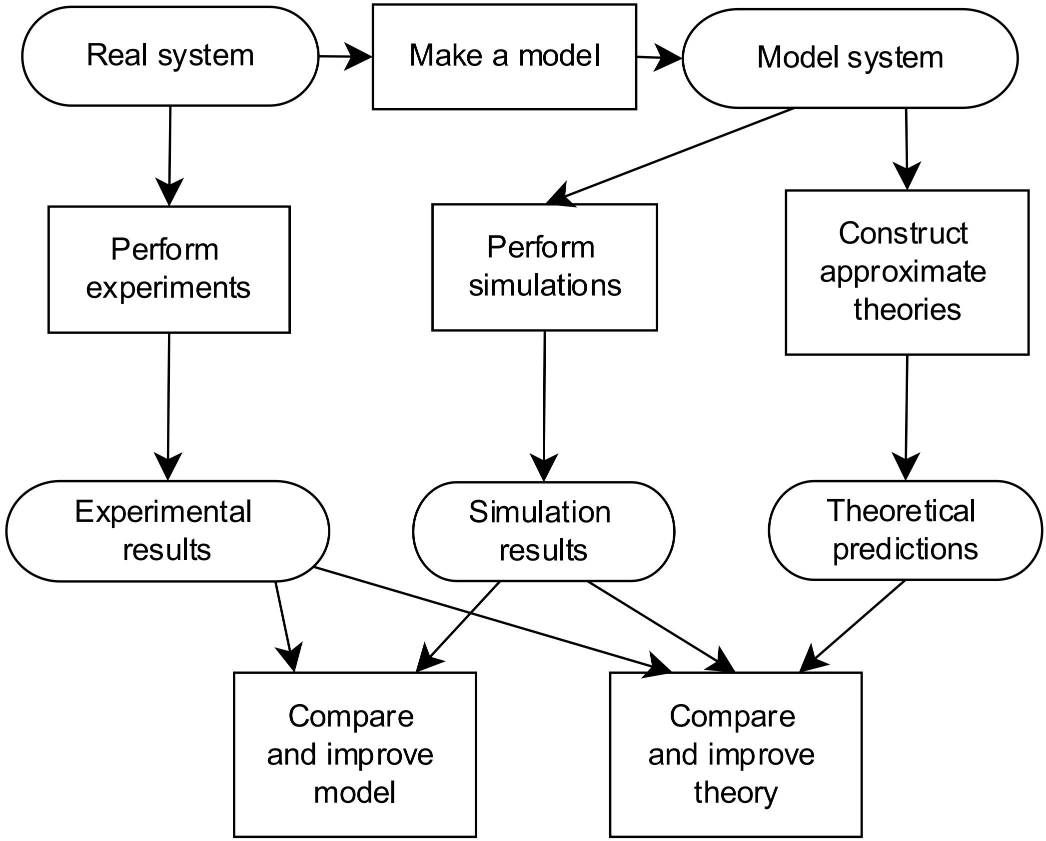
Austin J. Brockmeier, Ph.D.

Monday, March 20th, 2023

Probability and Spatial Statistics for Geospatial Systems and Data

1. Statistical analysis to understand the relationship of attributes in and across time and space (Lecture 6)
2. Generate data from random variables and processes through **computer simulation** (Lab 6&7)
3. Analyze spatial patterns in data (Lab 6&7)
4. Create **models** of data-generating processes (Lecture 7)

- **Lab 6**: generate & analyze spatial random variables
- **Lab 7**: spatial interpolation raster data



Outline

Probability theory and modeling for spatial data

- Joint, marginal, and conditional distributions
- Bayes rule
- Hypothesis testing
- Spatial lag/nearest neighbor as modeling
- Interpolation
- Voronoi diagram/clustering
- Linear/non-linear regression

Example Motivating Questions:

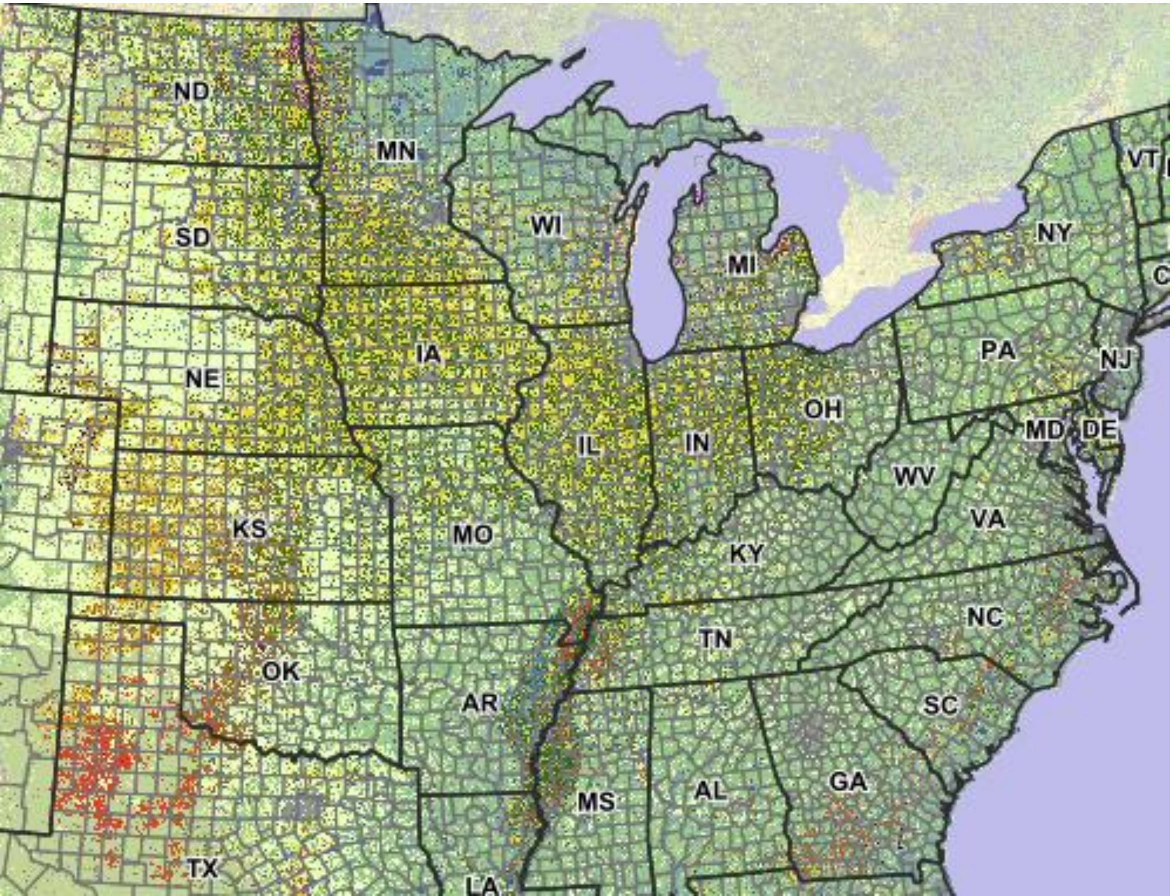
Abstract:

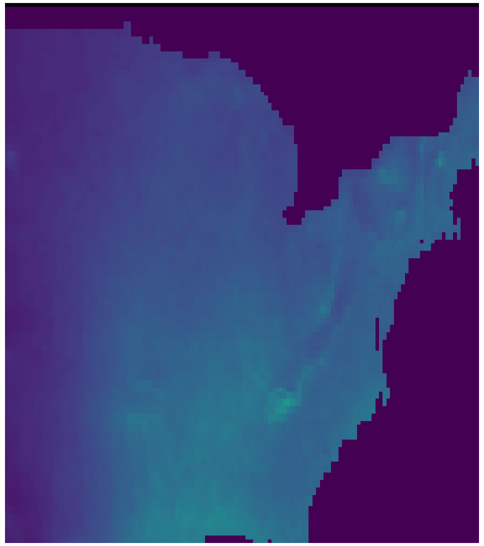
What is the most likely attribute category in a particular area?
What's the average and variation of attribute values in a particular area?
Do pairs of relatively nearby points have similarly valued features?
How does the value of one attribute co-relate with the category of another attribute?

Concrete:

- What is the most common crop raised in each section of land?
- What is the average rainfall in the month of July in Newark? How much does the rainfall vary per year?
- How related is the rainfall in Newark, Delaware to the rainfall at the beach in Lewes, Delaware?
- How dependent is the choice of crop on the rainfall across the US?

Example Motivating Questions:





- What is the most common crop raised in each section of land?
- How dependent is the choice of crop on the rainfall across the US?

A causal (based on knowledge) conditional probability:

$$\Pr\left(Y_{\text{crop}} = \text{'corn'} \middle| X_{\text{precip}} \in [10{,}22]\right)$$

Spatial autocorrelation

variable: attribute/measurement/feature

Applicable to **discrete objects** or **points in time or a spatial field**

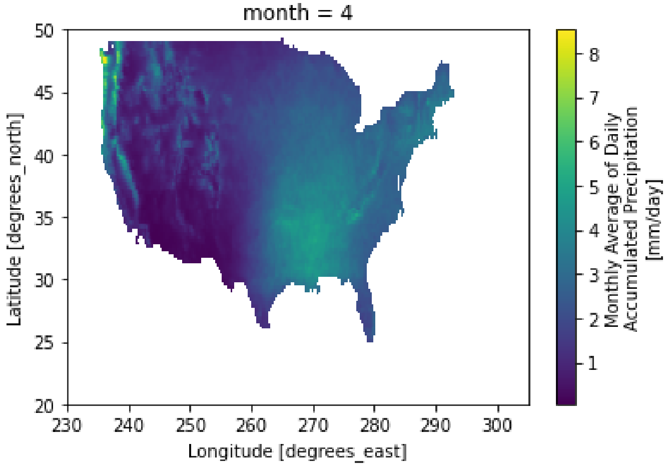
- Autocorrelation
  - Is there a (linear) relationship in the value of a variable for objects at relatively nearby locations l_1, l_2 ?

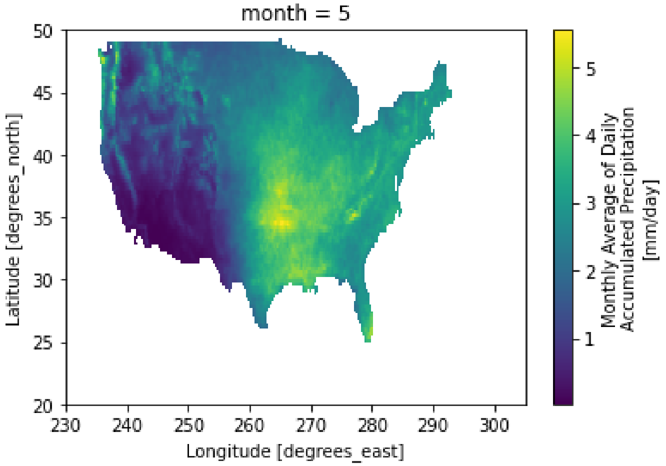
- Cross-correlation
  - Is there a (linear) relationship in the value of the variables for relatively nearby objects/points?

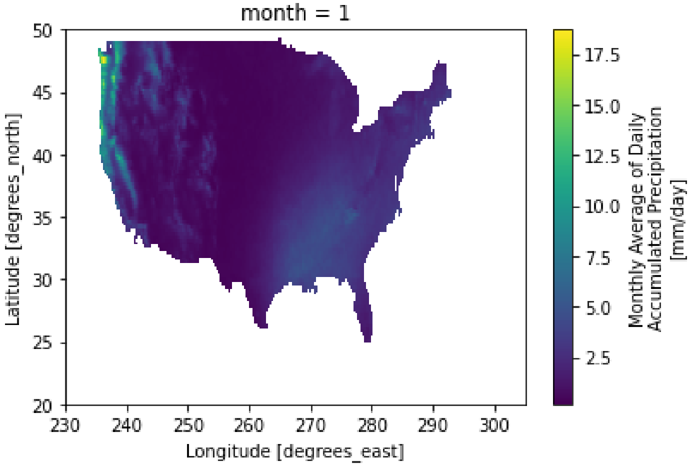
How does the strength (variance explained) of the relationship vary as a function of the distance between the objects/points?

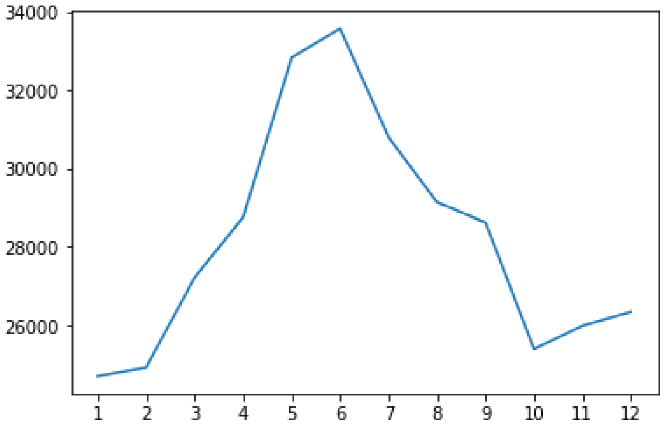
Conditioning and slicing

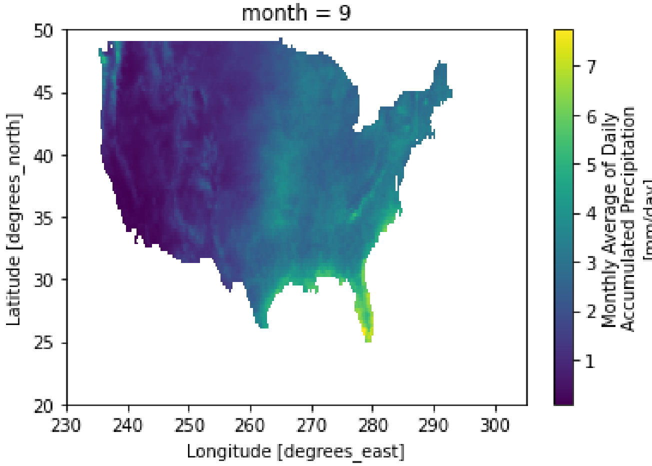
```
# Calculate the weighted average
precip_monthly = mon_precp_xr.groupby("time.month").mean(dim="time")
total_precip = np.array([precip_monthly['precip'][month].sum() for month in range(12)])
```





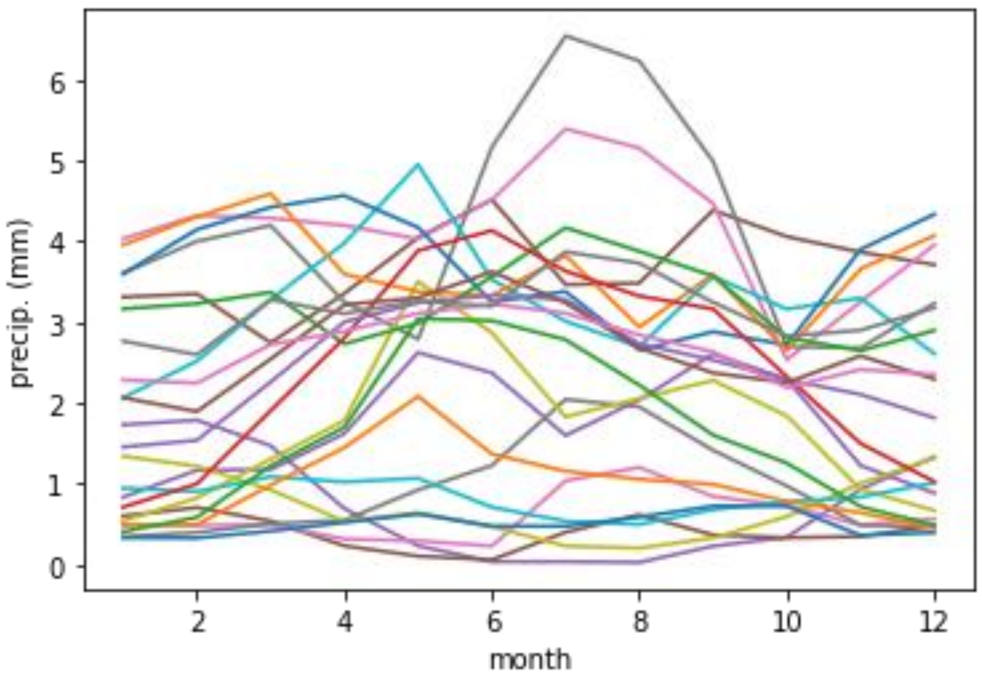


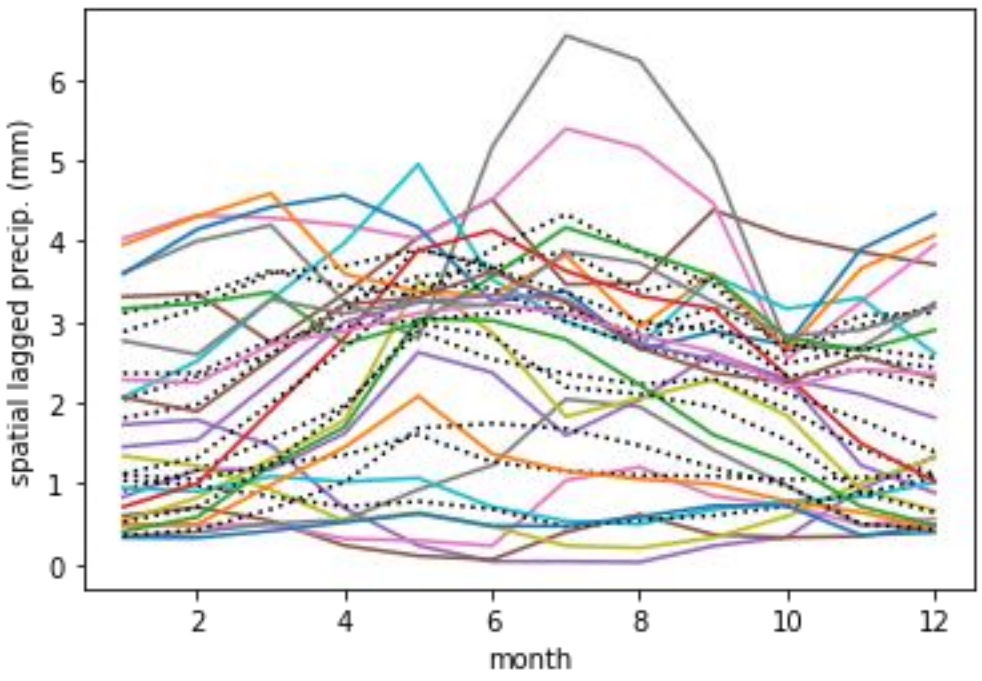




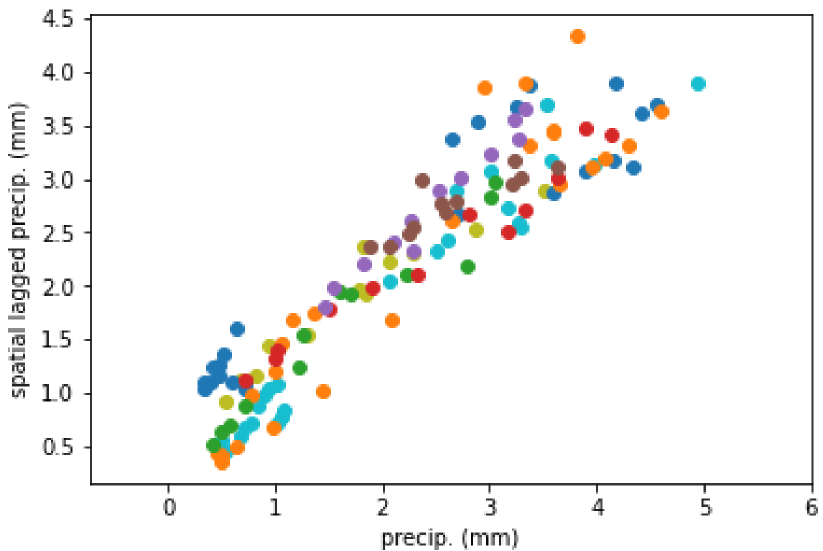
Spatial correlation

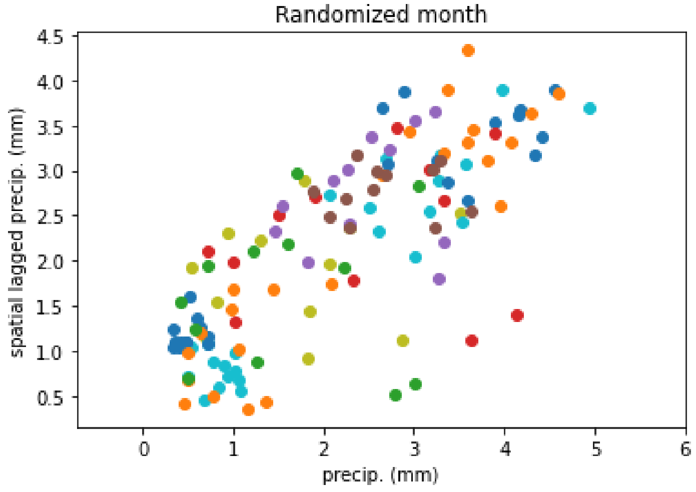
Neighbors at ±5 degrees



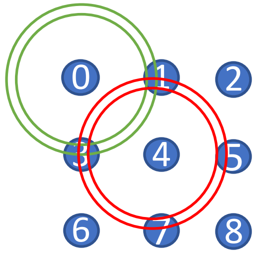


Spatial prediction = Average of 4 grid neighbors





Weight analysis with buffer neighborhood



Data:

$\{(x_i, y_i, \vec{\mathbf{z}}_i)\}_{i=1}^{n}$

- Row-normalized weight matrix of radius 1 buffer:

$$W \in \mathbb{R}^{n\times n}$$

$$\vec{\mathbf{z}}_i = \begin{bmatrix} z_i^{(1)} \\ \vdots \\ z_i^{(d)} \end{bmatrix} \in \mathbb{R}^d \text{, for } i = 1,\dots,n$$

Model:

$$\widehat{\vec{\mathbf{z}}_i} = f_W(\vec{\mathbf{z}}_i) = \sum_{j=1}^{n} W_{ij}\, \vec{\mathbf{z}}_j$$

Index	pt.x	pt.y
0	-1	1
1	0	1
2	1	1
3	-1	0
4	0	0
5	1	0
6	-1	-1
7	0	-1
8	1	-1

Index	0	1	2	3	4	5	6	7	8
0		$\frac{1}{2}$		$\frac{1}{2}$					
1	$\frac{1}{3}$		$\frac{1}{3}$		$\frac{1}{3}$				
2		$\frac{1}{2}$				$\frac{1}{2}$			
3	$\frac{1}{3}$				$\frac{1}{3}$		$\frac{1}{3}$		
4		$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$	
5			$\frac{1}{3}$		$\frac{1}{3}$				$\frac{1}{3}$
6				$\frac{1}{2}$				$\frac{1}{2}$	
7					$\frac{1}{3}$		$\frac{1}{3}$		$\frac{1}{3}$
8						$\frac{1}{2}$		$\frac{1}{2}$	

k-nearest neighbor

Data:
$\{(x_i, y_i, \vec{\mathbf{z}}_i)\}_{i=1}^{n}$

- Distance: $d_{ij} = d([x_i, y_i], [x_j, y_j])$
- How many closer? $o_{ij} = |\{t \neq i : d_{it} < d_{ij}\}|$
- k-Neighborhood

$$\mathcal{N}_k(i) = \{j \neq i : o_{ij} < k\}$$

- Compute weight matrix: $W \in \mathbb{R}^{n \times n}$

$$W'_{ij} = \begin{cases} 1, & j \in \mathcal{N}_k(i) \\ 0, & \text{otherwise} \end{cases}$$

Row normalized:

$$W_{ij} = \frac{W'_{ij}}{\sum_{k=1}^{n} W'_{ik}}$$

Conditional distribution

$f_{X|Y=1}$

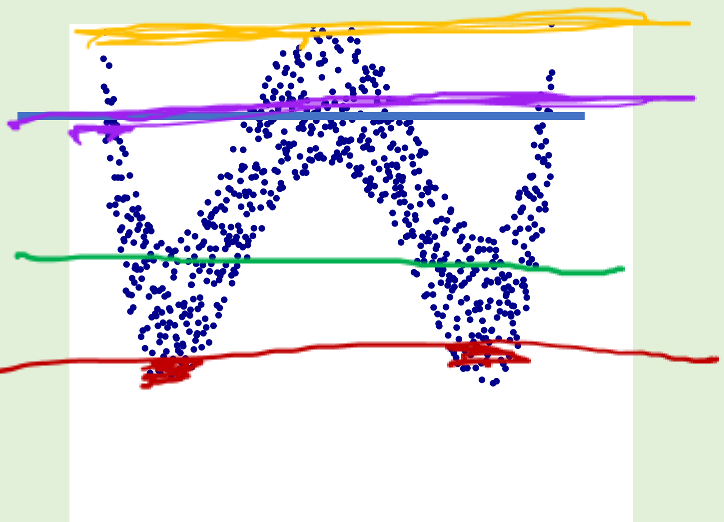
$f_{X|Y=y}(x)$

Example 4: **A sample of points from the joint distribution of *X* and *Y***

The marginal is shown. What does the conditional density of X given Y=0.8 look like?

$y = 0.8$

0.2



$f_X(x)$

By Ainali - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3141713

$f_{X|Y=0.8}$

$f_{X|Y=0.6}$

$f_{X|Y=0.2}$

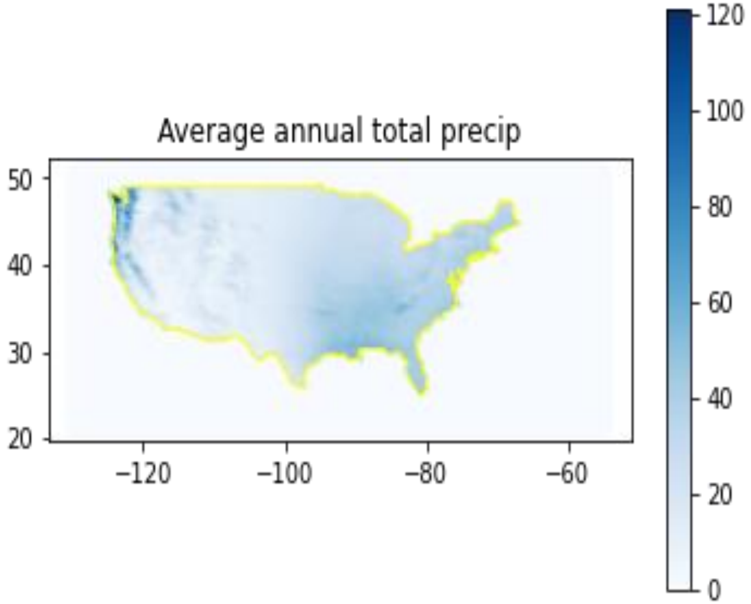
- Under the hypothesis that the two are independent $p_{XY}(x, y) = p_X(x)p_Y(y)$

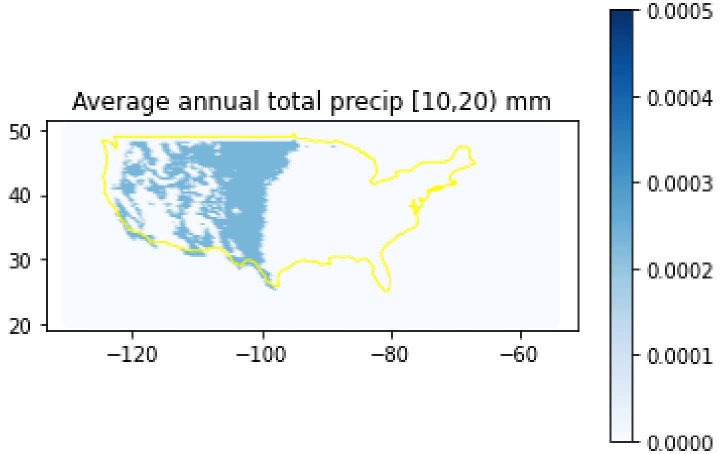
Does the data indicate dependence?

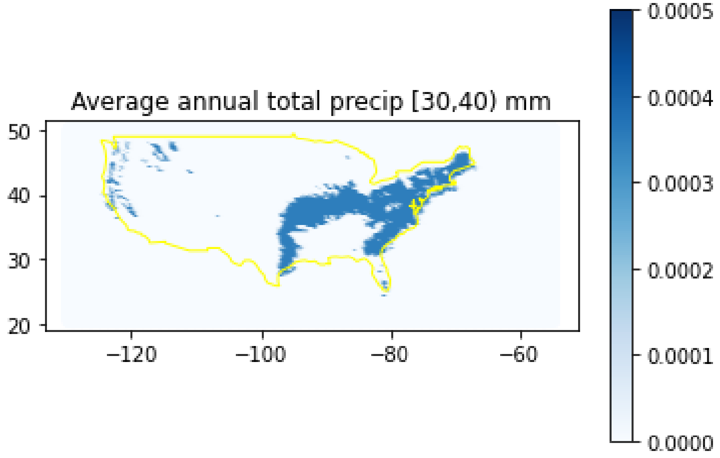
Conditional distributions

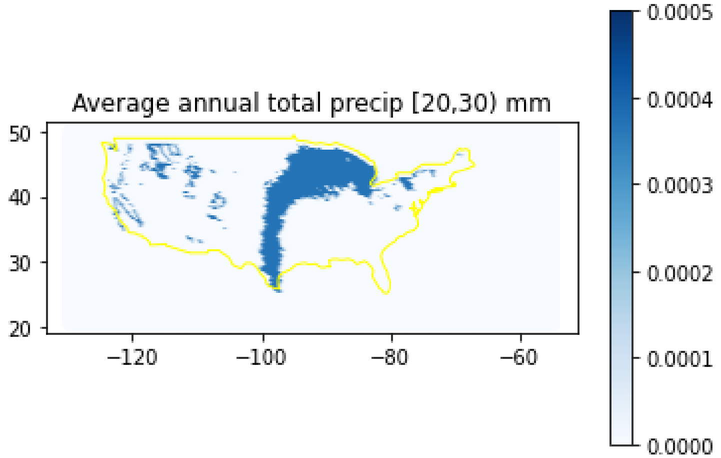
- Conditional PMF

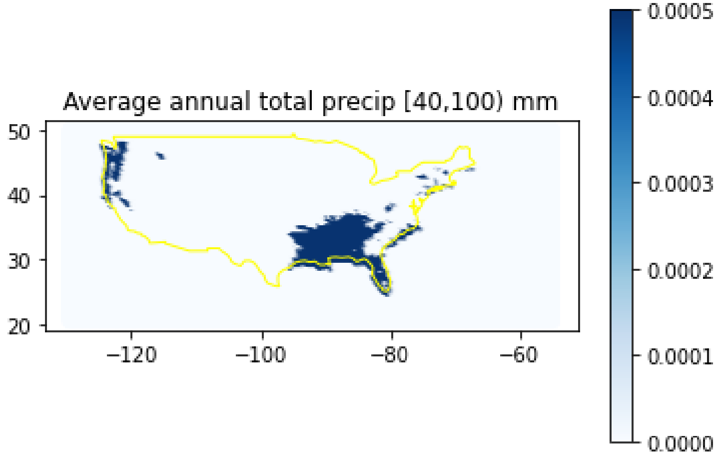
  Pr(X=x|Precip. is in range)









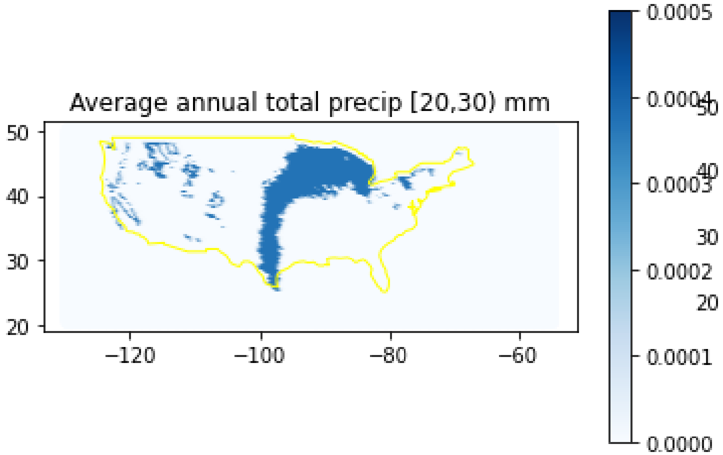


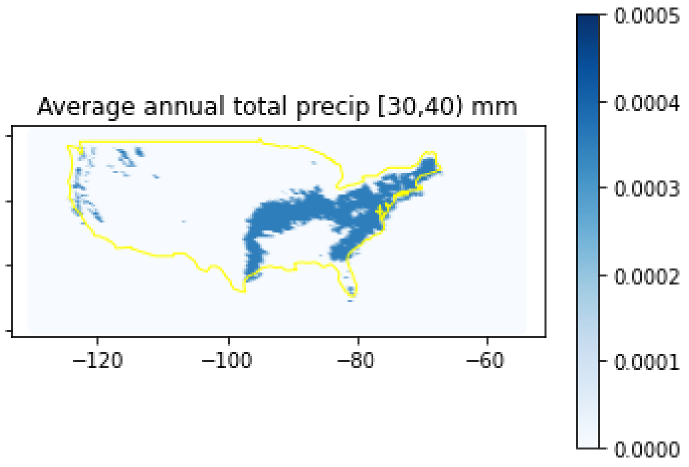
Bayes theorem

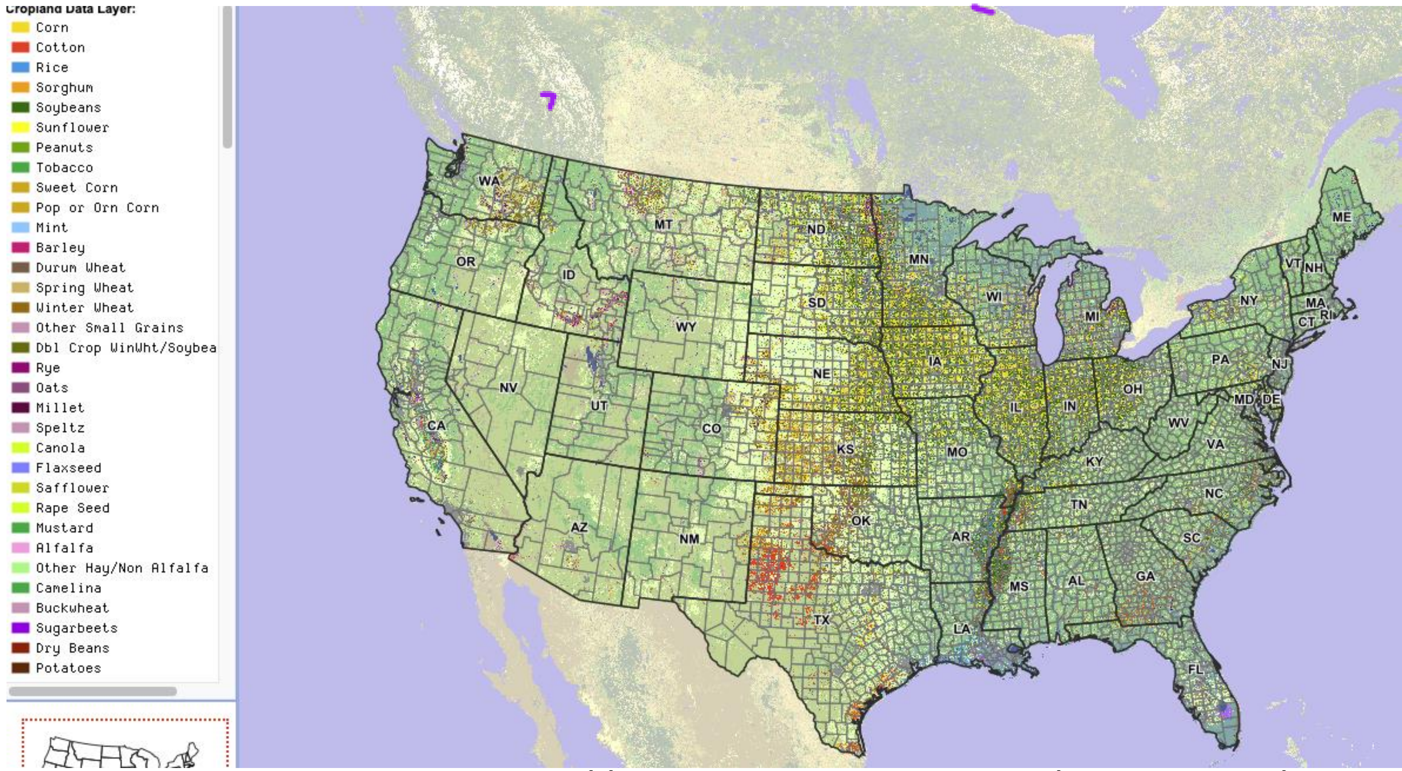
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

P(Corn | Precip. in range)=

	Precip [20,30) Yes	No	
Corn is most planted	0.1	0.1	0.2
Not corn	0.1	0.7	0.8
	0.2	0.8	1







https://nassgeodata.gmu.edu/CropScape/

Testing Dependence

- Under the hypothesis that the two are independent $P(A,B) = P(A)P(B)$

Does the data indicate dependence?

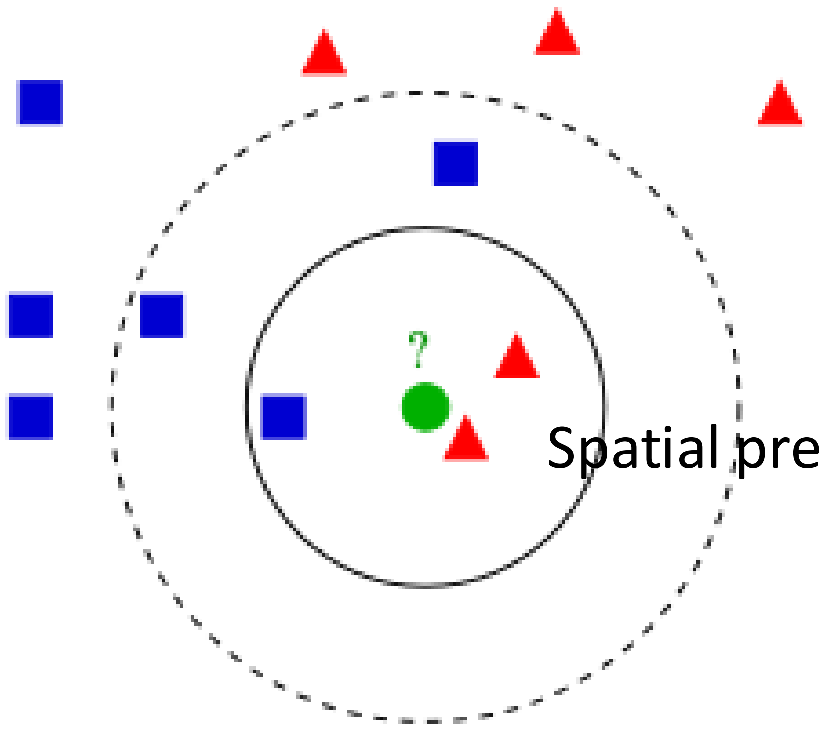
	Precip [20,30) Yes	No	
Corn is most planted	0.1	0.1	0.2
Not corn	0.1	0.7	0.8
	0.2	0.8	1

https://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}$$

$$= N\sum_{i,j} p_{i\cdot}p_{\cdot j}\left(\frac{(O_{i,j}/N)-p_{i\cdot}p_{\cdot j}}{p_{i\cdot}p_{\cdot j}}\right)^2$$

Spatial correlation for categorical attributes

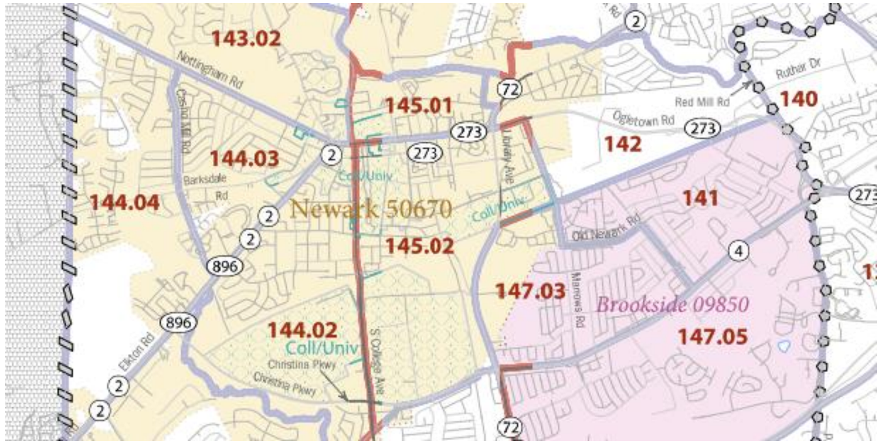


Spatial prediction = Majority vote of *k* neighbors

This is a model/hypothesis of reality.
How to test if the model is valid?

US census tracts designed to help understand the relationship between attributes

Census tracts are optimized groupings for statistical analysis



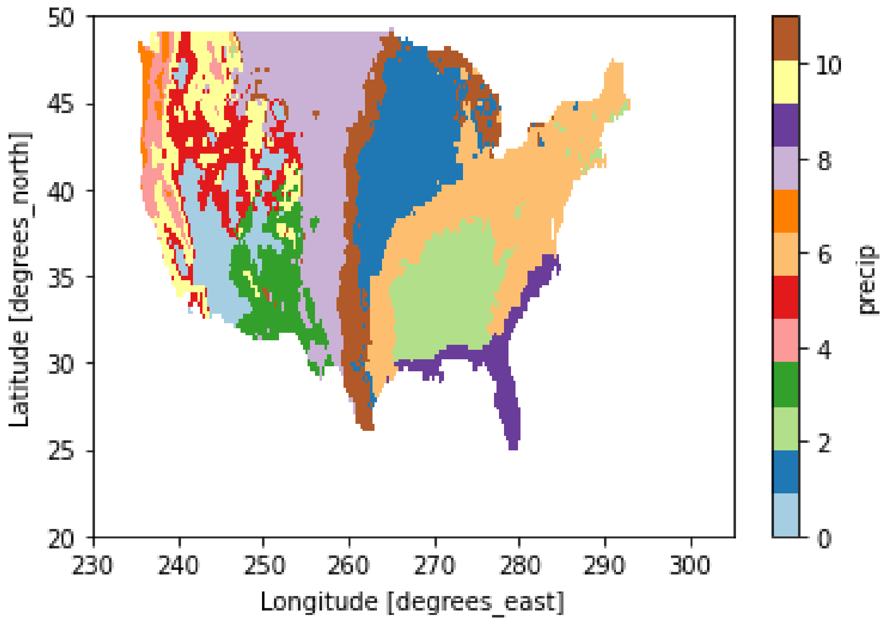
https://www2.census.gov/geo/maps/DC2020/PL20/st10_de/censustract_ma 10003_new_castle/DC20CT_C10003.pdf

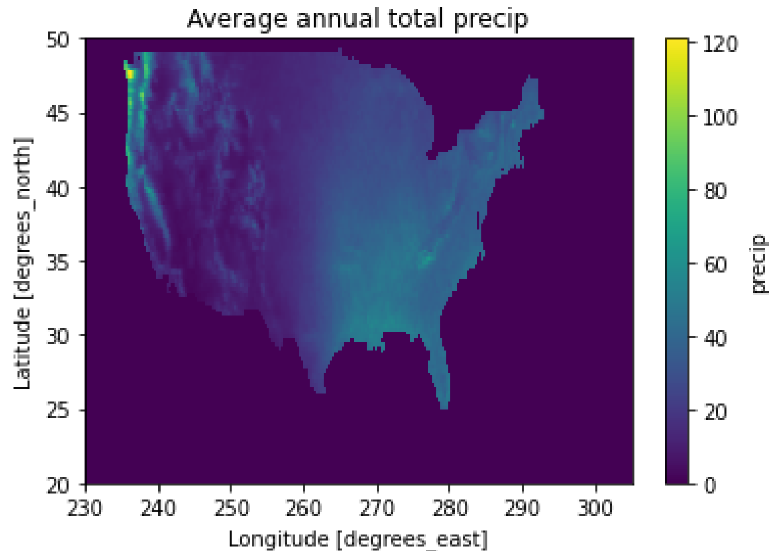
https://blog.rtwilson.com/wp-content/uploads/2012/01/SnowMap_Points.png

Nearest neighborhoods form Voronoi diagrams

Grouping data by attributes w/o geometry

Clusters assigned by pattern across months





Source image.

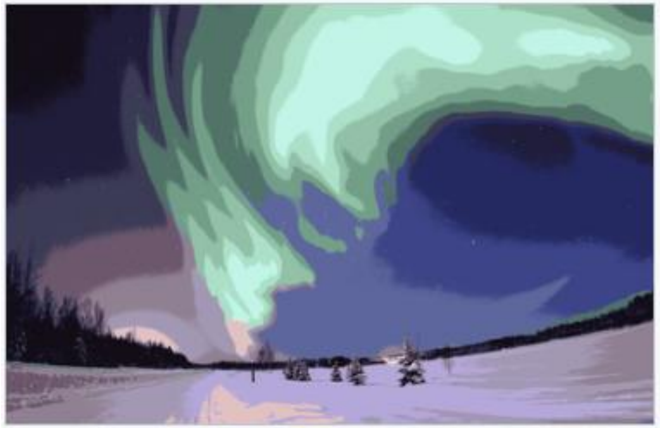
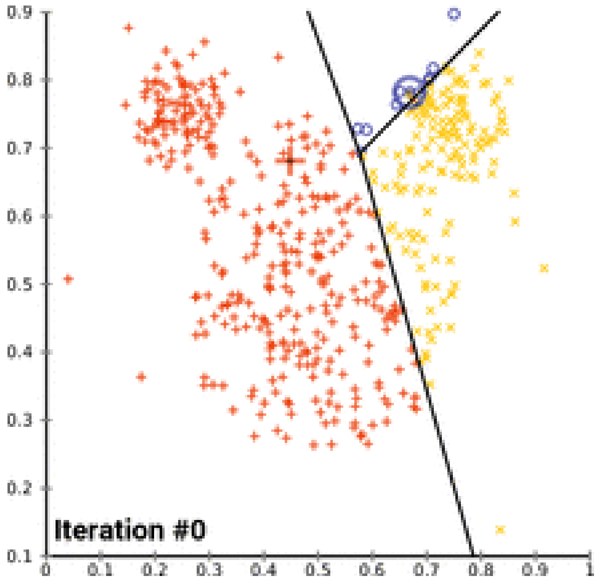


Image after running *k*-means with *k = 16*. Note that a common technique to improve performance for large images is to downsample the image, compute the clusters, and then reassign the values to the larger image if necessary.

K-means clustering

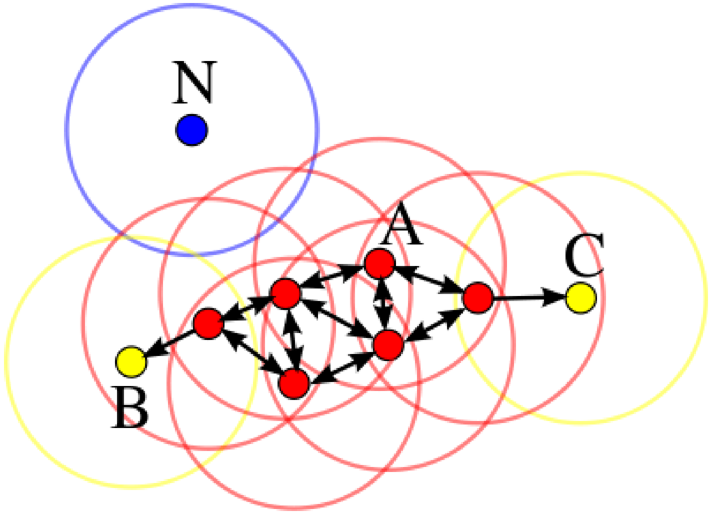


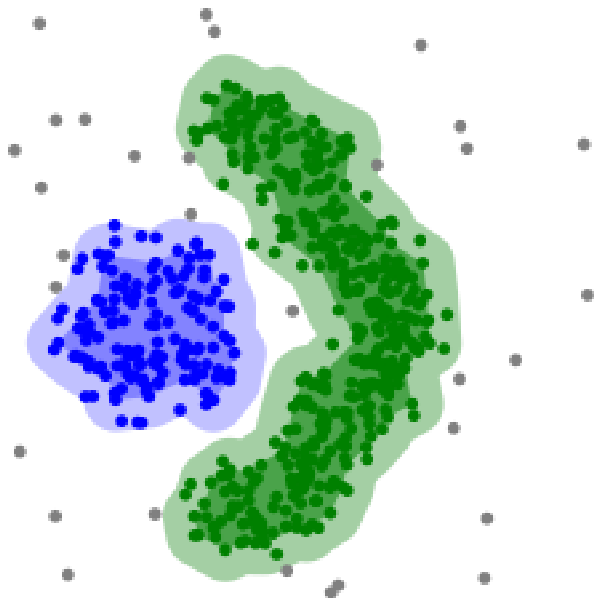
0. Assign some cluster centers ‘centroids’ randomly
1. Assign each point to the nearest centroid
2. Find the mean of all points assigned to each cluster’s centroid; this mean becomes the new centroid of the cluster
3. Repeat steps 1&2 until convergence

Grouping points by density-based clustering

DBSCAN (radius/buffer size, minimum number in each)

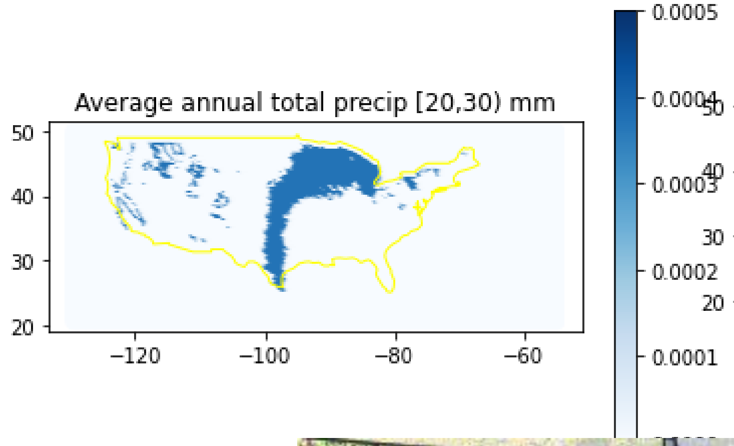
1. Find points in dense areas—those with **<u>enough close</u>** neighbors
2. Find chains of these dense-area points, include neighbors that do not meet density requirement
3. Exclude other points as noise

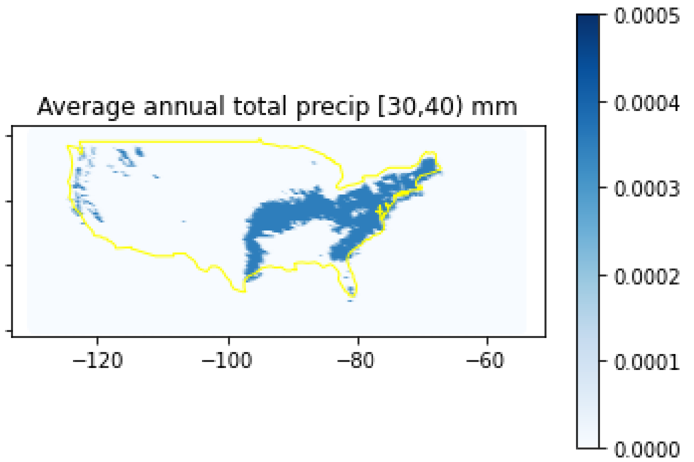


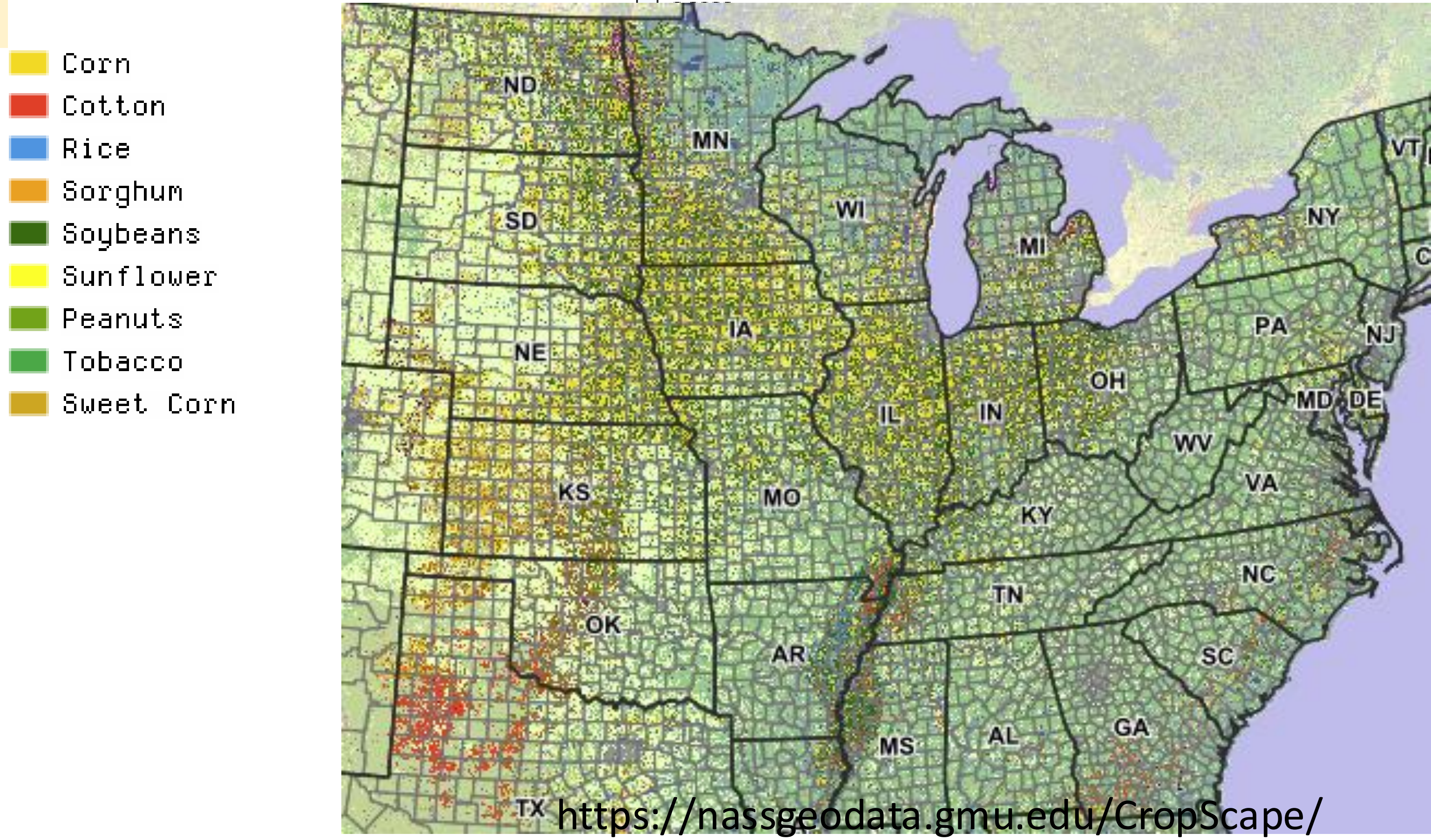


Modeling

What other factors should be used to explain crop production?

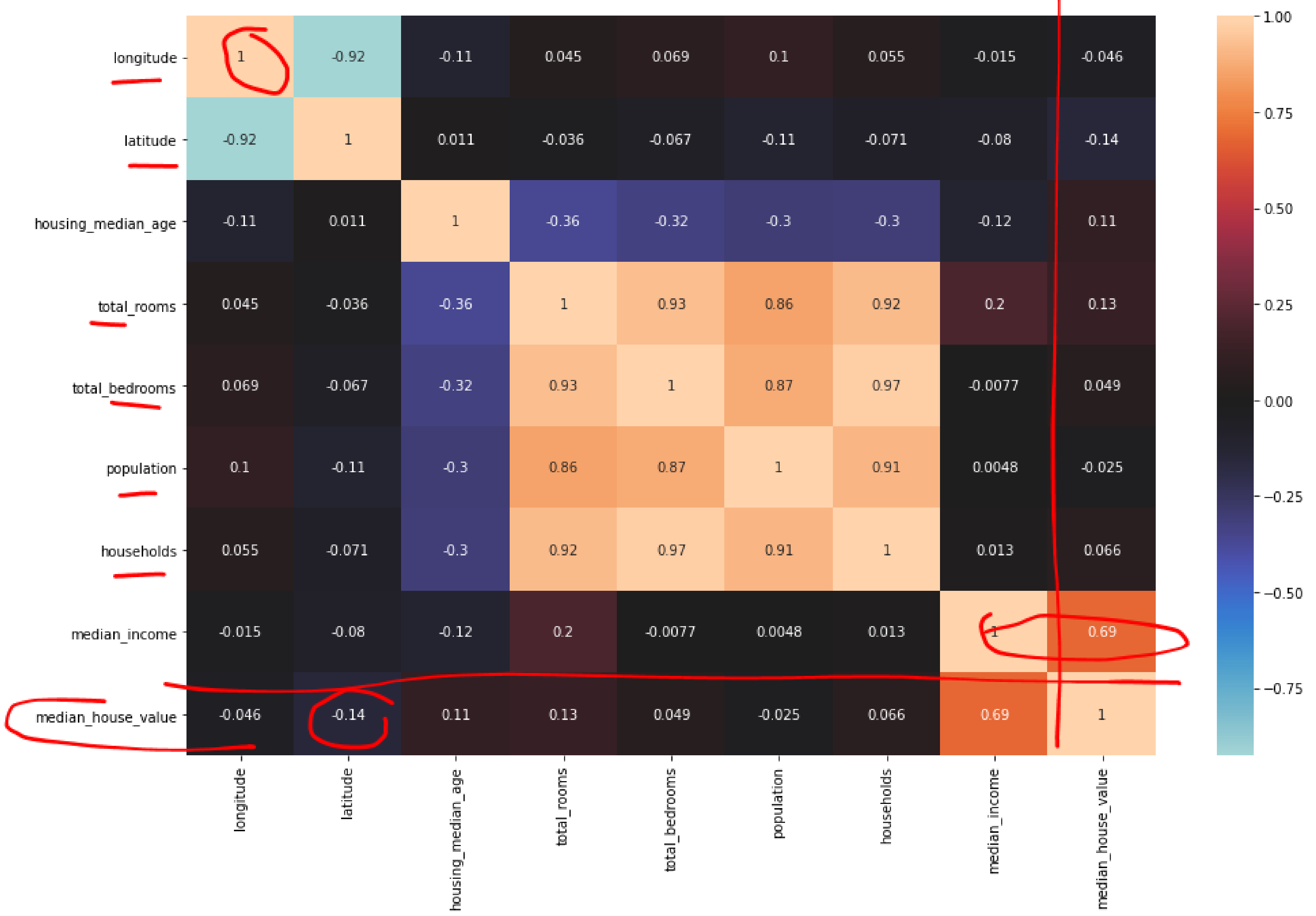






https://nassgeodata.gmu.edu/CropScape/

Correlation in attributes for CA house prices



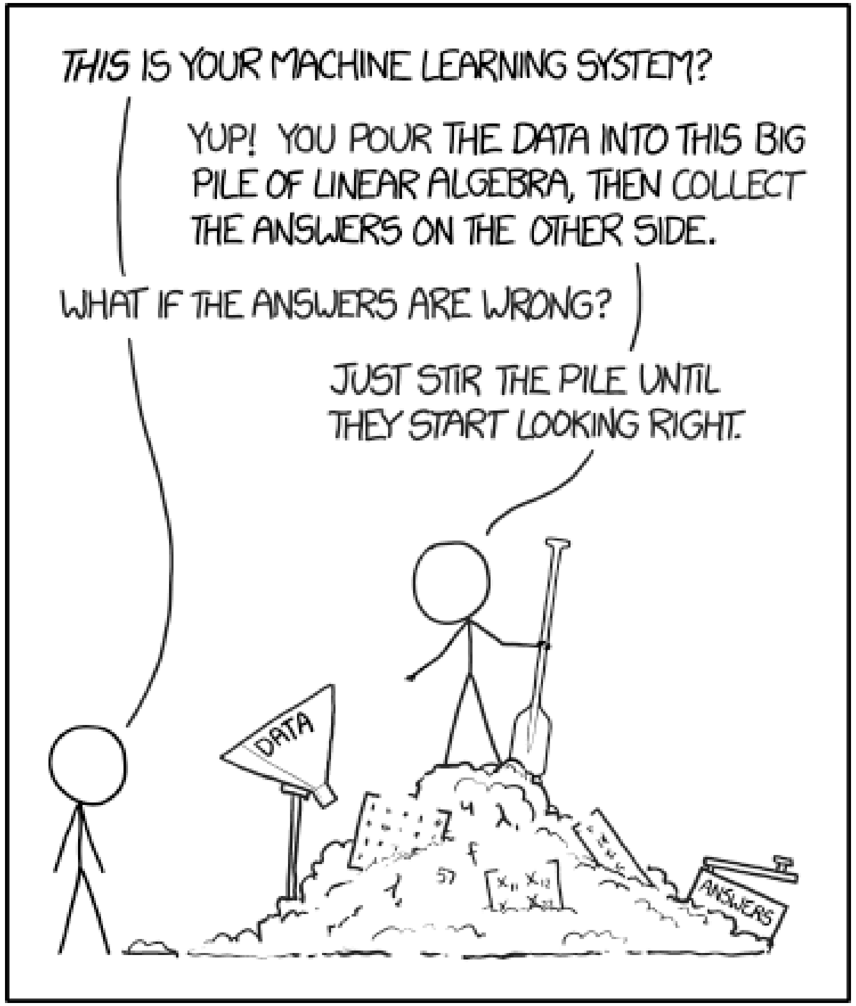
Combining these features together

		$\vec{x}_1$	$\vec{x}_2$	…	$\vec{x}_n$
housing_median_age	$x^{(1)}$				
total_rooms	$x^{(2)}$				
total_bedrooms	$x^{(3)}$				
	$\vdots$				
population	$x^{(d)}$				
median_house_value	y	y_1	y_2	…	y_n
		House 1	House 2		House n

$$\hat{y} = w \cdot \vec{\boldsymbol{x}} + b = \sum_{i=1}^{d} x^{(i)} w_i + b$$

Machine learning

- Linear Regression
- Kernel Regression
- Gaussian Process Regression/Kriging



Linear model for regression

	$\vec{x}_1$	$\vec{x}_2$	…	$\vec{x}_n$
$x^{(1)}$				
$x^{(2)}$				
$x^{(3)}$				
$\vdots$				
$x^{(d)}$				
y	y_1	y_2	…	y_n

Data:

$\{(\vec{\boldsymbol{x}}_i, y_i)\}_{i=1}^{n}$

$$\vec{\boldsymbol{x}}_i = \begin{bmatrix} x_i^{(1)} \\ \vdots \\ x_i^{(d)} \end{bmatrix} \in \mathbb{R}^d, \quad y_i \in \mathbb{R} \ \text{ for } i = 1,\ldots,n$$

Model:

$$f_\theta(\vec{\boldsymbol{x}}) = w \cdot \vec{\boldsymbol{x}} + b = \hat{y}, \qquad \theta = [w_1, \ldots, w_d, b] \in \mathbb{R}^{d+1}$$

Loss: $J(\theta) = \sum_{i=1}^{n} \frac{1}{n} |y_i - f_\theta(x_i)|^2$

Optimal solution: $\theta^* = \underset{\theta}{\operatorname{argmin}}\, J(\theta)$

Linear model for regression

$f_{\theta^*}(\vec{\boldsymbol{x}}\ ) = w^* \cdot (\vec{\boldsymbol{x}} - \bar{\vec{\boldsymbol{x}}})\ + b^*$

$\theta^*\ = [w_1^*, \dots, w_d^*, b] \in \mathbb{R}^{d+1}$

$b^*\ = \sum_{i=1}^{n} \frac{1}{n} y_i\ ,$

$w^*\ = \Sigma^{-1} \vec{\boldsymbol{p}}$

$\Sigma\ \ = \mathrm{Cov}(\vec{\boldsymbol{x}})$

$$\Sigma_{kl} = \mathrm{Cov}(x^{(k)}, x^{(l)})$$

$\vec{\boldsymbol{p}} = \mathrm{Cov}(\vec{\boldsymbol{x}}, y)$

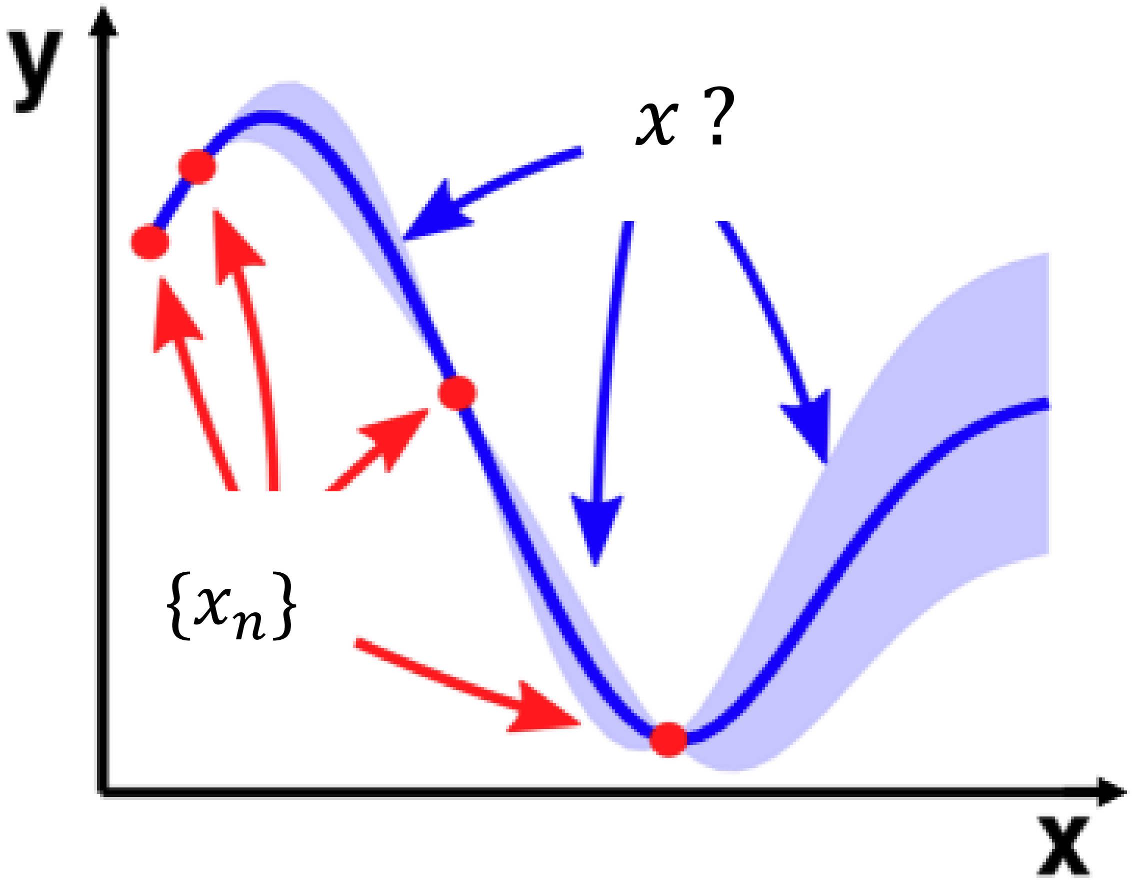
$$p_k = \mathrm{Cov}\left(x^{(k)}, y\right)$$

$$= n^{-1} \sum_i y_i (\, x_i^{(k)} - \bar{x}^{(k)})$$

Σ	$\boldsymbol{x}^{(1)}$	$\boldsymbol{x}^{(2)}$	…	$\boldsymbol{x}^{(d)}$
$\mathbf{x}^{(1)}$	σ_1^2	$\rho\sigma_2\sigma_1$		$\rho\sigma_d\sigma_2$
$\boldsymbol{x}^{(2)}$	$\rho\sigma_1\sigma_2$	σ_2^2		
$\vdots$			$\ddots$	
$\boldsymbol{x}^{(d)}$	$\rho\sigma_1\sigma_d$			σ_d^2

Local models

- k-nearest neighbor
- Decision trees, random forests, gradient boosting
- Neural networks
- Kernel ridge regression
- Gaussian process

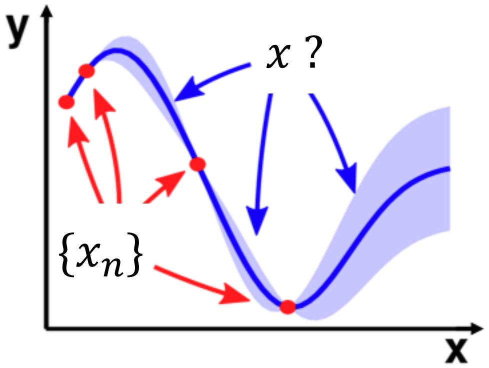


Phase 1. Fit relationship
Phase 2: Find x that gives a specific y with high confidence (near seen data) and fits constraints!

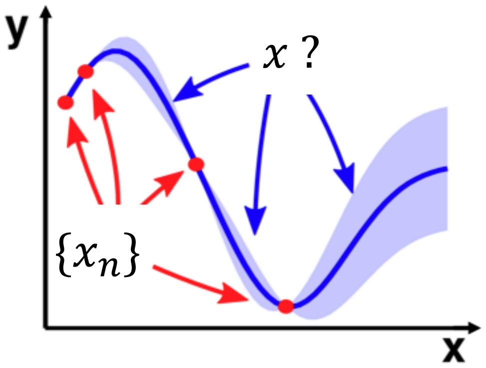
- Kernel regression
  - Advanced by Prof. Grace Wahba at UW-Madison

$$E[Y|x,\{(x_i,y_i)\}_{i=1}^{n}] = \bar{f}(x) = [\kappa(x,x_1),\ldots,\kappa(x,x_n)]\mathbf{K}^{-1}\vec{\boldsymbol{y}} = \boldsymbol{K}\vec{\boldsymbol{\alpha}}$$

```
krr.fit(X,y).predict(x)
```



- Gaussian process



http://www.infinitecuriosity.org/vizgp/
https://distill.pub/2019/visual-exploration-gaussian-processes
http://www.tmpl.fi/gp/
https://gaussianprocess.org/

The predicted value at x is normal with mean, $\mathcal{N}(f(x), \sigma_x^2)$

$$\sigma_x^2 = \mathrm{cov}\big(f(x), f(x)\big) = \kappa(x, x) - [\kappa(x, x_1), \dots, \kappa(x, x_N)]\mathbf{K}^{-1}[\kappa(x, x_1), \dots, \kappa(x, x_N)]$$

Upcoming in Lab 7

- Quiz!
- Bayes rule: joint and conditional
- Spatial correlation
- Clustering