地理信息系统与遥感应用

第十五讲 遥感图像分类

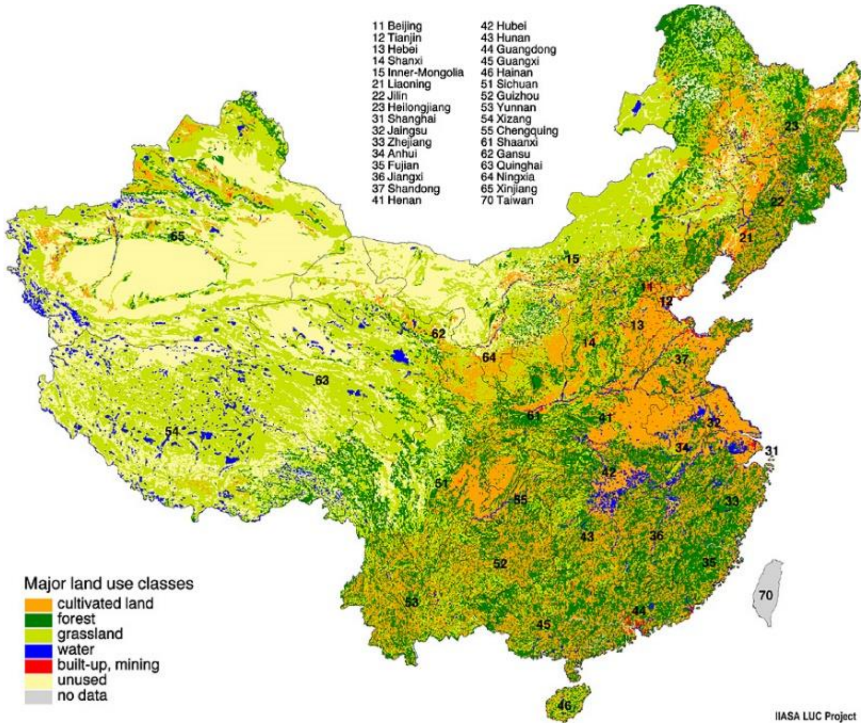
南方科技大学 · 环境科学与工程学院

田 勇

2018年12月17日

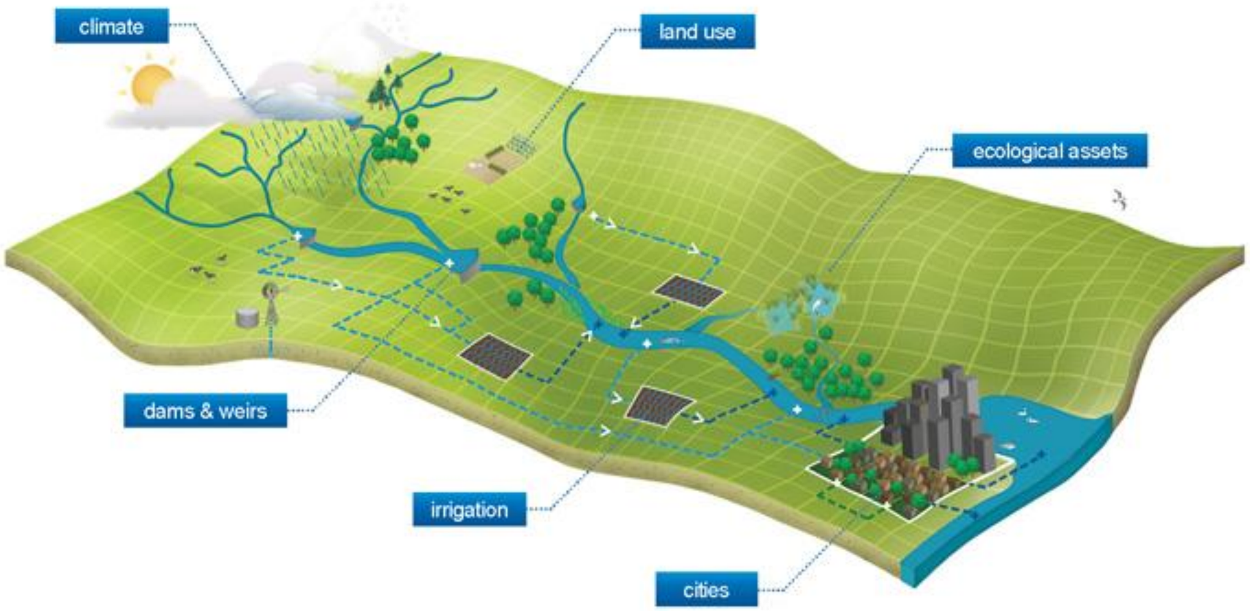
Image Classification

- Why classify?
- Make sense of a landscape
  - – Place landscape into categories (classes)
    - Forest, Agriculture, Water, etc



Example Uses

- Provide context
  - Landscape planning or assessment
  - Research projects
- Drive models
  - Hydrological Models
  - Meteorology Models
  - Biodiversity Models



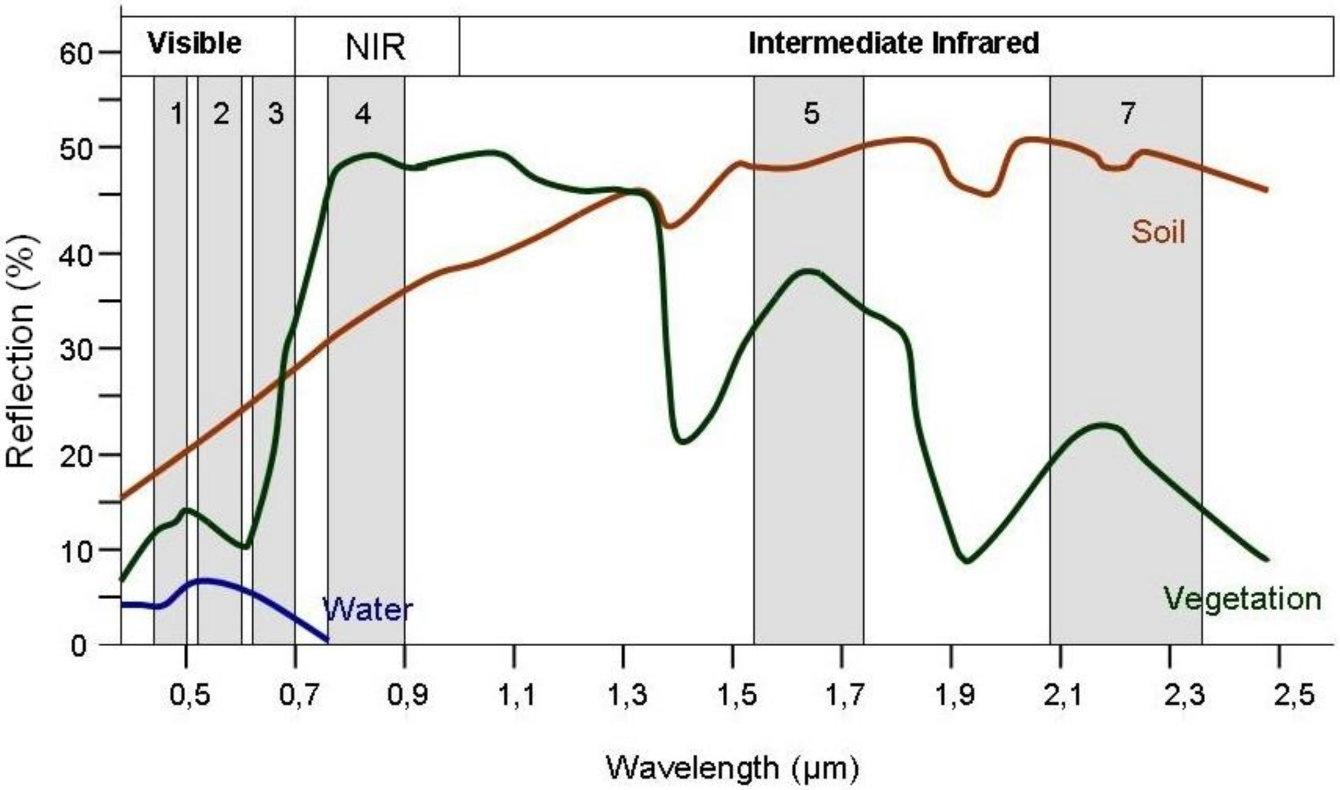
Classification

TODAY’S PLAN

- Basic strategy for classifying remotely sensed images using spectral information
- Supervised Classification
    - Spectral Angle Mapper Method
    - K Nearest Neighbor (KNN)
    - Maximum Likelihood
    - Principle Component Analysis (PCA)
- Unsupervised Classification
- Lab 15

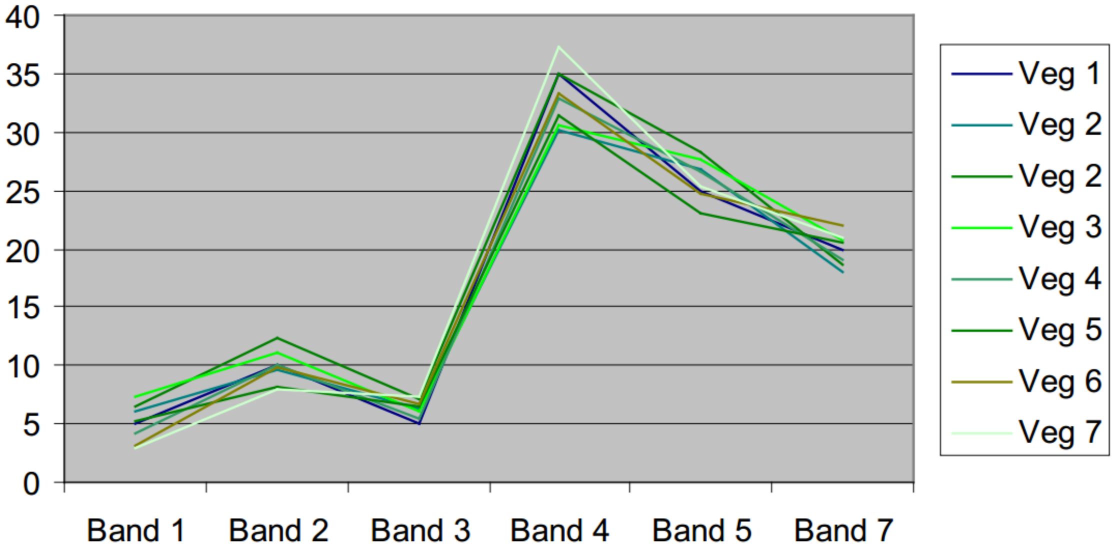
Basic Strategy: How do you do it?

- Use radiometric properties of remote sensor
- Different objects have different spectral signatures
- All “Vegetation” pixels would have exactly the same spectral signature
- Then we could just say that any pixel in an image with that signature was vegetation
- We’d do the same for soil, etc. and end up with a map of classes



SUSTech | Environment of Science

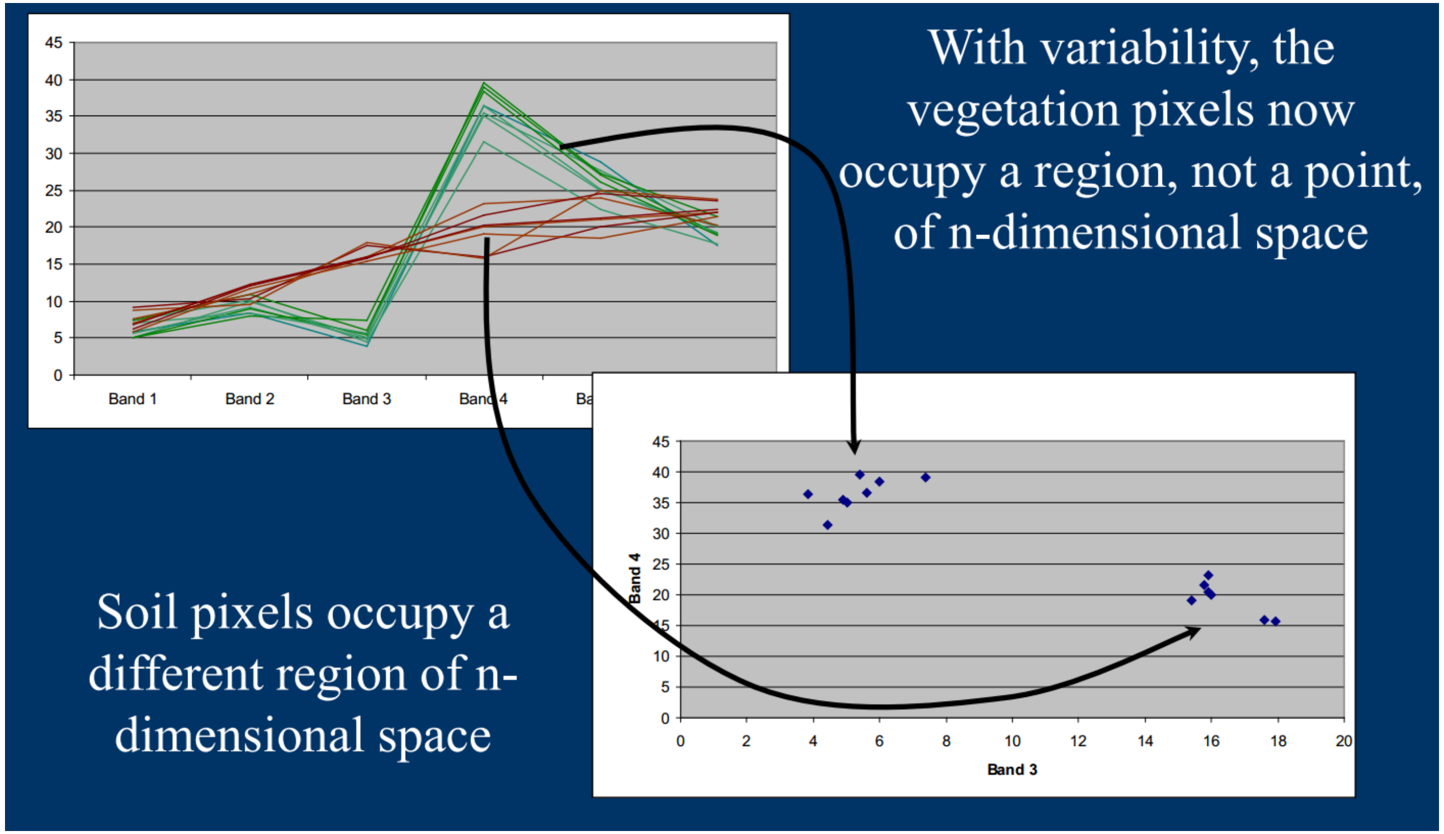
But in reality, that isn’t the case. Looking at several pixels with vegetation, you’d see variety in spectral signatures.



The Classification Trick: Deal with variability

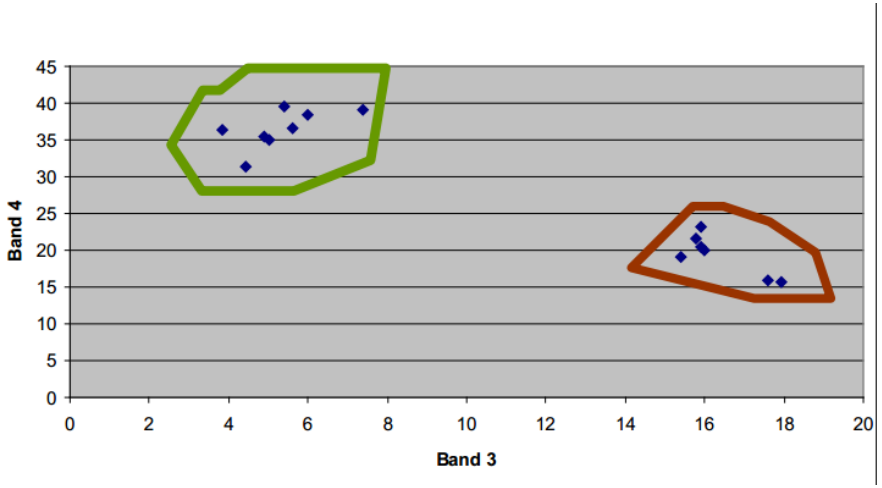
- Different ways of dealing with the ariability lead to different ways of classifying images
- To talk about this, we need to look at spectral signatures a little differently

Basic Strategy: Dealing with variability



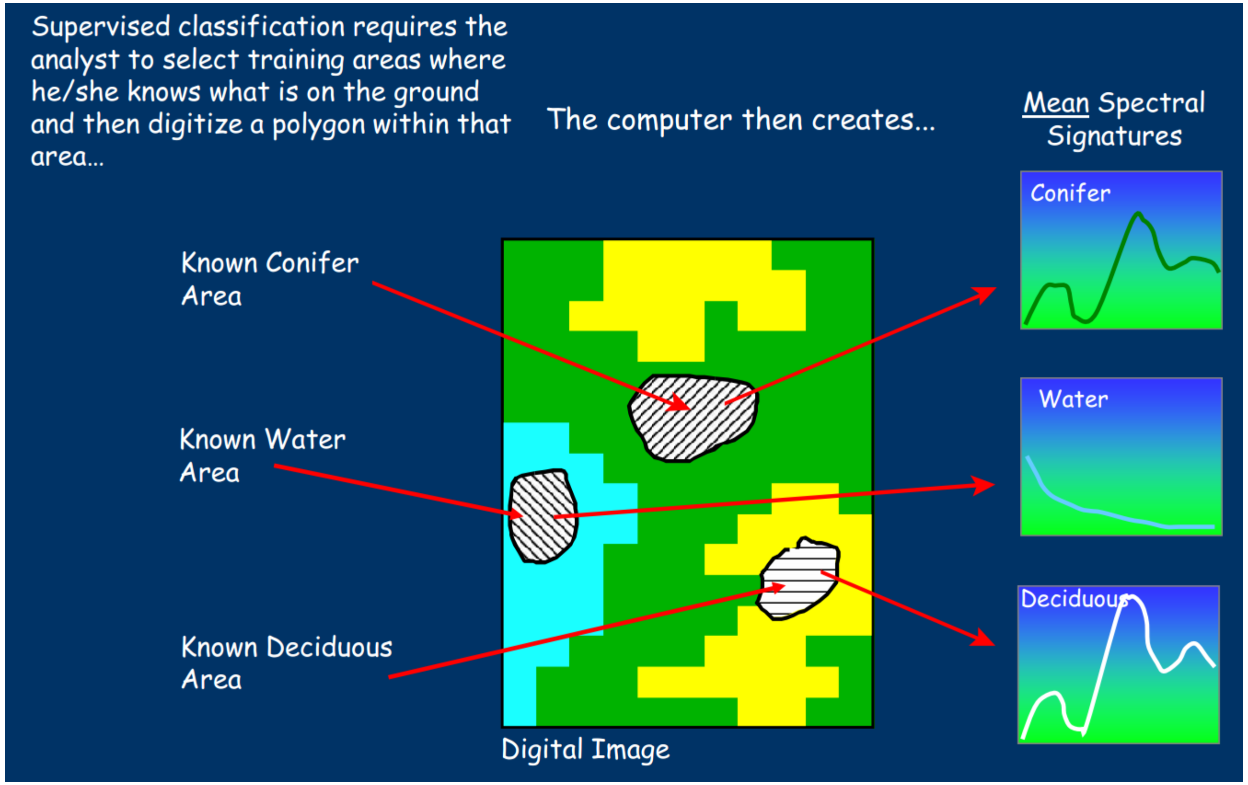
Classification:

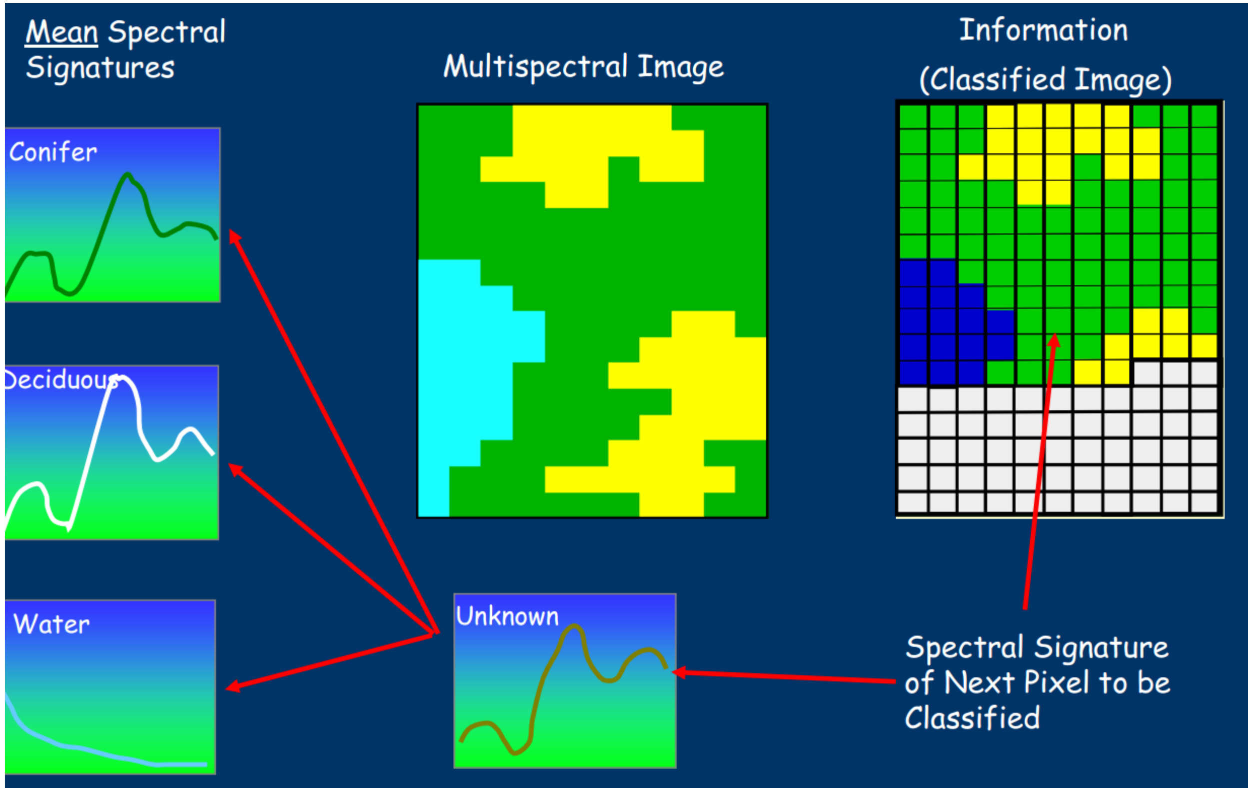
- Delineate boundaries of classes in n-dimensional space
- Assign class names to pixels using those boundaries



Classification Strategies

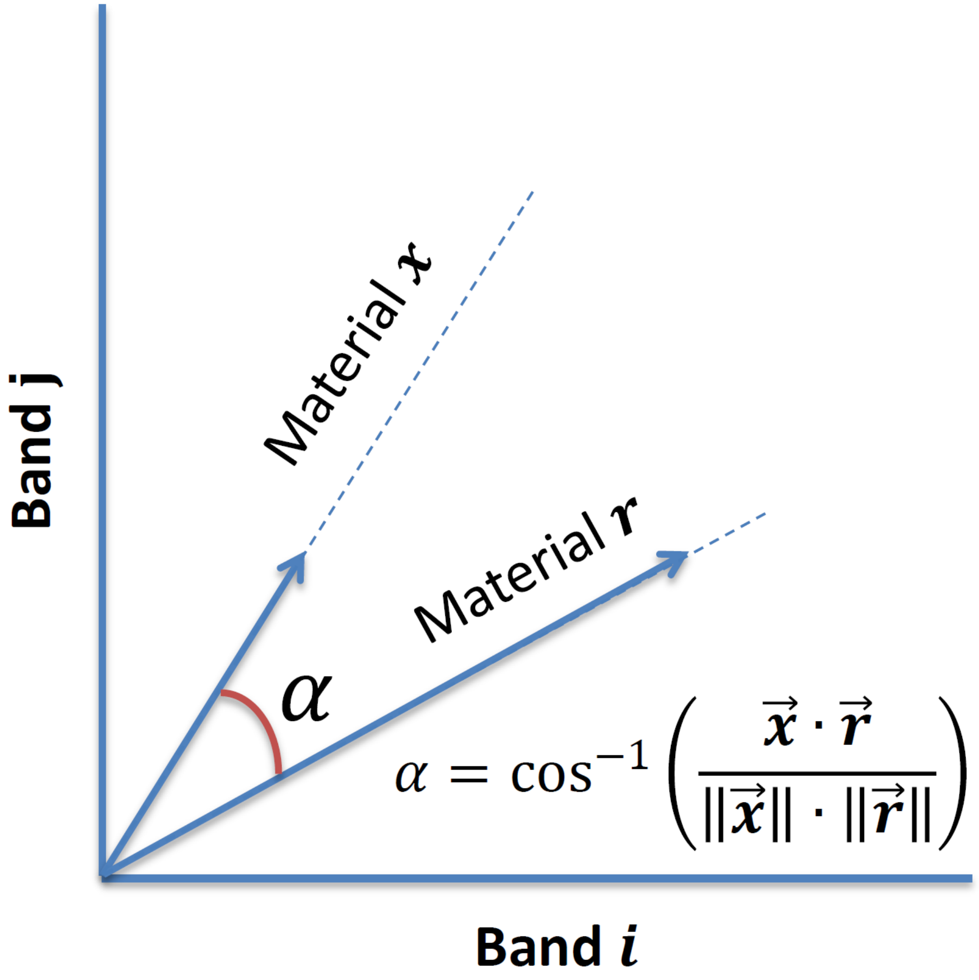
- Two basic strategies
  - Supervised classification
    - We impose our perceptions on the spectral data
  - Unsupervised classification
    - Spectral data imposes constraints on our interpretation





Supervised classification

(1) Supervised Classification: Spectral Angle Mapper

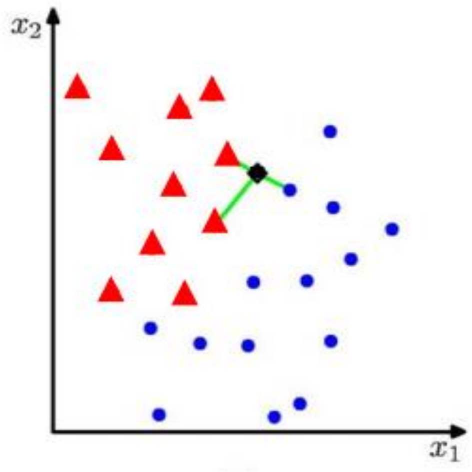


(2) **Supervised Classification: K Nearest Neighbor (KNN)**

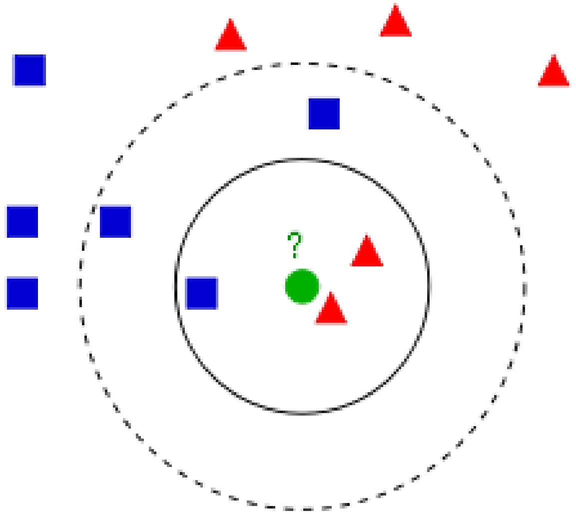
Algorithm

- For each test point, x, to be classified, find the K nearest samples in the training data
- Classify the point, x, according to the majority vote of their class labels

e.g. K = 3

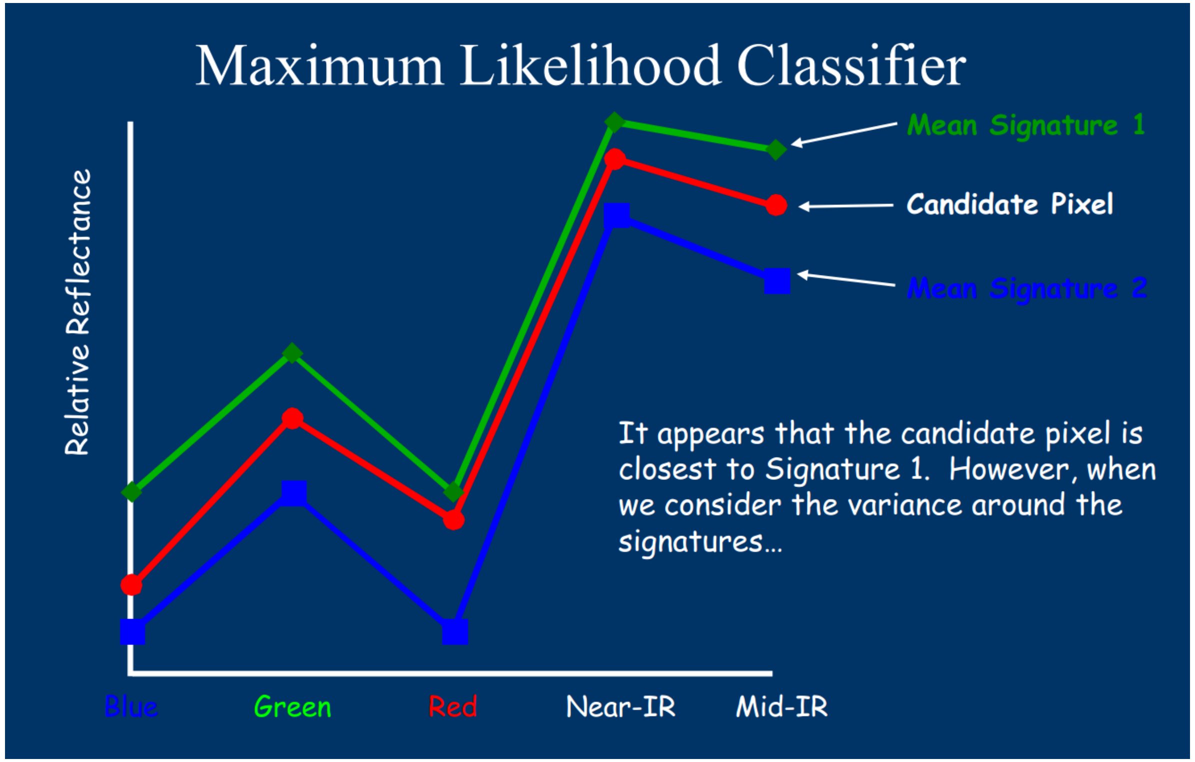


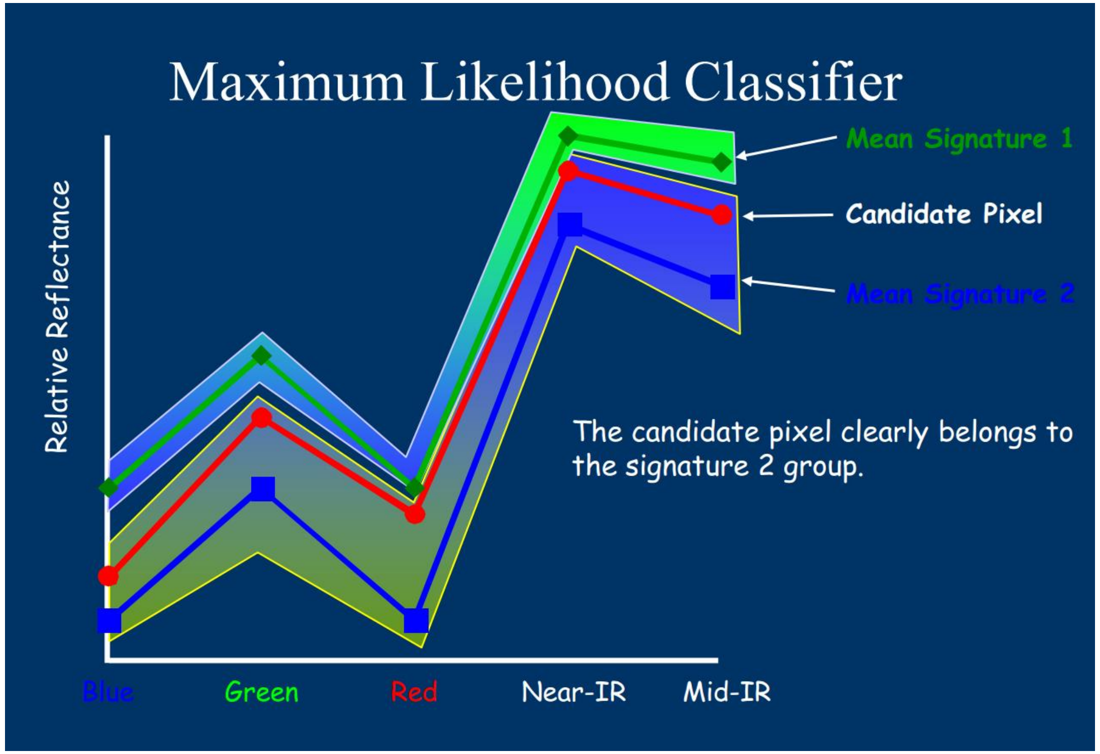
- applicable to multi-class case



(3) Supervised Classification: Maximum Likelihood

- Assume multivariate normal distributions of pixels within classes
- For each class, build a discriminant function
  - – For each pixel in the image, this function calculates the probability that the pixel is a member of that class
  - – Takes into account mean and covariance of training set
- Each pixel is assigned to the class for which it has the highest probability of membership





感性认识：

现在我们有两个盒子：甲和乙，每个里面都装了100个球，其中甲中装了95个红球，5个黑球，乙中装了60个红球，40个黑球，现在有人从盒子里面取出了一个球，发现是红球，然后让你猜：“他是从哪个盒子里面取出来的？”

基本原理：

假设有两个事件A，B，我们通过先验的知识，知道A发生的条件下，x也发生的概率是P(x|A)； B发生的条件下，x也发生的概率是P(X|B)，那么，现在有一个事件x发生了，我们能否判断这个事件x是在A条件下，还是在B条件下发生的可能性大些呢？也就是要求出P(A|x)和P(B|x)哪一个最大？对分类问题而言，哪一个概率大，我们就说x属于哪一类

贝叶斯公式：

$$P(A|B)=P(B|A)*P(A)/P(B)$$

从贝叶斯公式中我们可以看到，求概率P(A|B)的问题转化成了求P(B|A)，P(A)和P(B)的问题.通常，我们事先能知道P(A)，P(B)；或者是P(A)和P(B)在分类问题中是公共的项，可以约去；再或是他们的差异可以忽略不计，所以，要P(A|B)最大，也就是要P(B|A)最大！而对于P(B|A)，我们可以从事先已经发生的事件中，通过统计等数学方法计算得到

根据贝叶斯公式，构建下面的目标函数：

$$L(x) = p(Ci|x) = p(x|Ci) * p(Ci)/p(x)$$

$L(x)$ 目标函数

$p(Ci|x)$ 已知x事件发生了，那么它属于Ci的概率

$p(x|Ci)$ 类别Ci 对应的x的概率值

$p(Ci)$ 类别Ci发生的概率

$p(x)$ 事件x发生的概率

N维空间的最大似然分类法原理：

对于多维空间的中的变量，其正态分布的概率密度函数

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |S_k|}} \exp\left[-\frac{1}{2}(x-\mu_k)^T S_k^{-1}(x-\mu_k)\right]$$

$f(x)$ 概率密度

n 特征维数，对遥感图像分类来说， 就是波段的个数

x n维空间中的一个向量

μ_k 也是一个n向量，它是由每一维特征的均值组成的一个向量

S_k n维特征向量之间的协方差矩阵

$|S_k|$ 协方差矩阵的行列式

假设x服从正态分布的时候，则x的概率密度函数：

$$p(x|Ci) = \frac{1}{\sqrt{(2\pi)^n |S_{Ci}|}} \exp\left[-\frac{1}{2}(x-\mu_{Ci})^T S_{Ci}^{-1}(x-\mu_{Ci})\right]$$

$p(Ci)$ 通常可以根据已知条件计算得到，或者其区别可以忽略不计

$p(x)$ 是公共项，所以不用考虑其具体的值

$p(x)$ 是公共项，去掉之后目标函数就变为：

$$L'(x) = p(Ci|x) = p(x|Ci) * p(Ci)$$

对目标函数取对数：

$$Ln(L'(x)) = Ln(p(Ci|x)) = Ln(p(x|Ci)p(Ci)) = Ln(p(x|Ci)) + Ln(p(Ci))$$

$$= Ln\left(\frac{1}{\sqrt{(2\pi)^n |S_{Ci}|}}\right) - \frac{1}{2}(x-\mu_{Ci})^T S_{Ci}^{-1}(x-\mu_{Ci}) + Ln(p(Ci))$$

$$\mathbf{C_x} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

(4) **Supervised Classification: Principal Component Analysis**

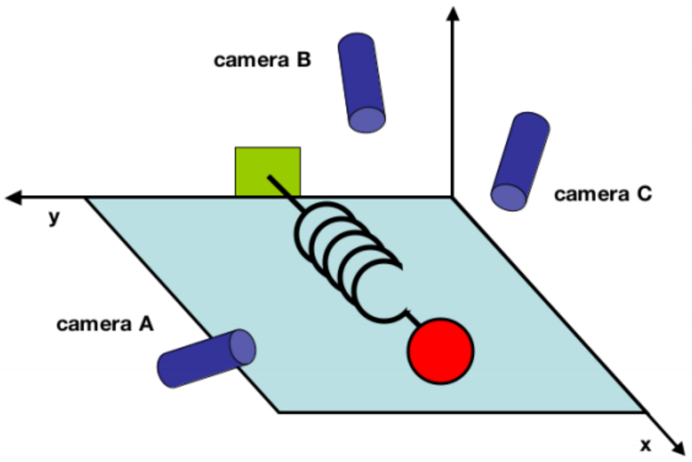


Figure 1: A diagram of the toy example.

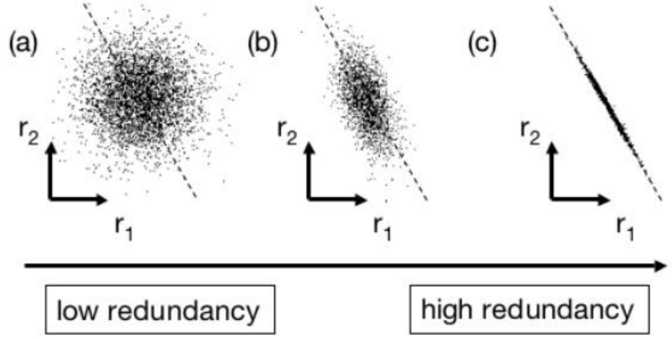


Figure 3: A spectrum of possible redundancies in data from the two separate recordings r_1 and r_2 (e.g. x_A, y_B). The best-fit line $r_2 = kr_1$ is indicated by the dashed line.

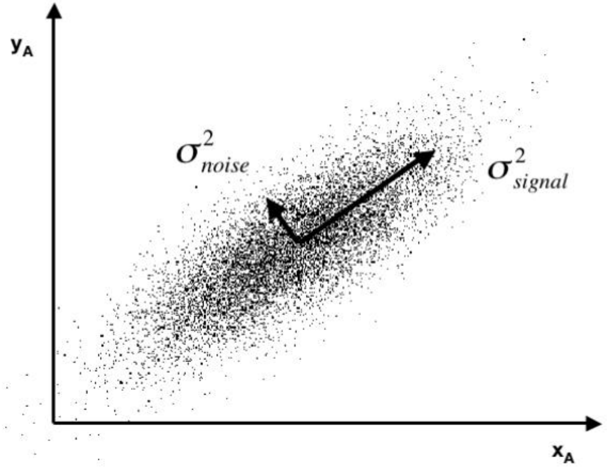
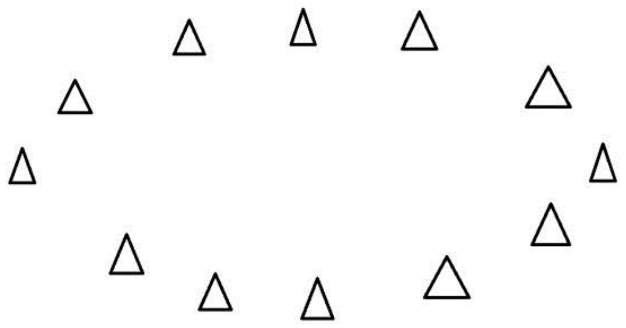


Figure 2: A simulated plot of (x_A, y_A) for camera A. The signal and noise variances σ^2_{signal} and σ^2_{noise} are graphically represented.

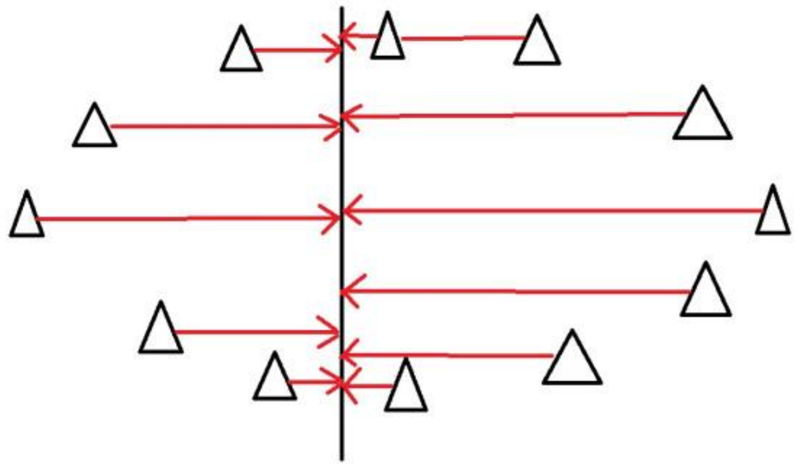
Motivation of Principal Component Analysis : A Toy Example

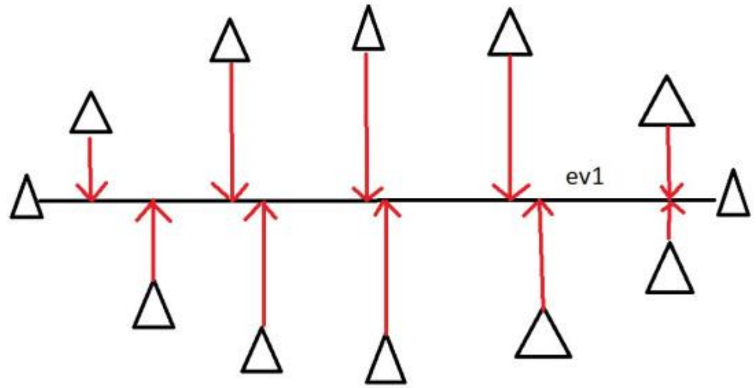
What is PCA?



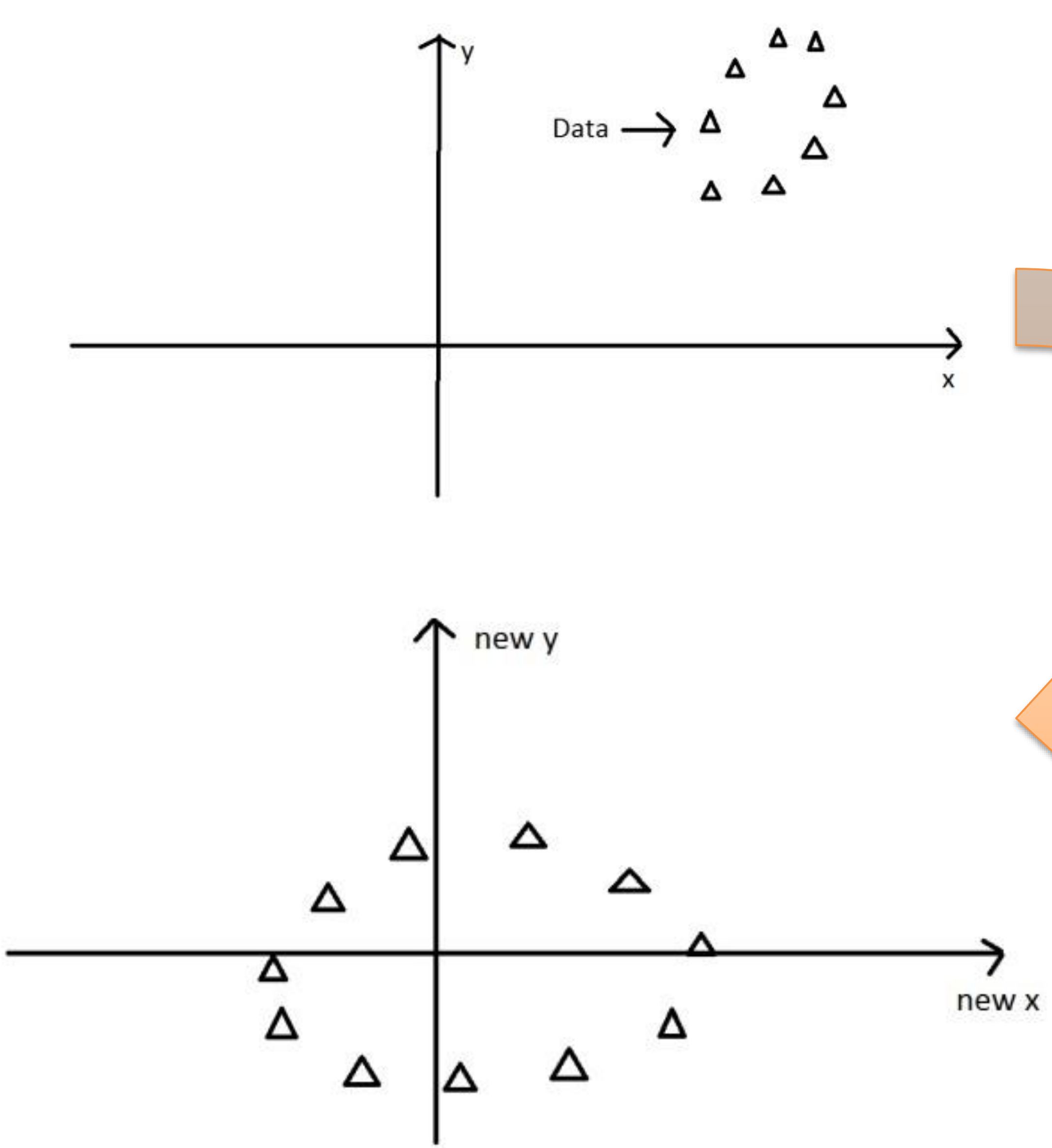
It is often useful to measure data in terms of its principal components rather than on a normal x-y axis. So what are principal components then? They're the underlying structure in the data. **They are the directions where there is the most variance**, the directions where the data is most spread out.

Find the straight line where the data is most spread out when projected onto it.

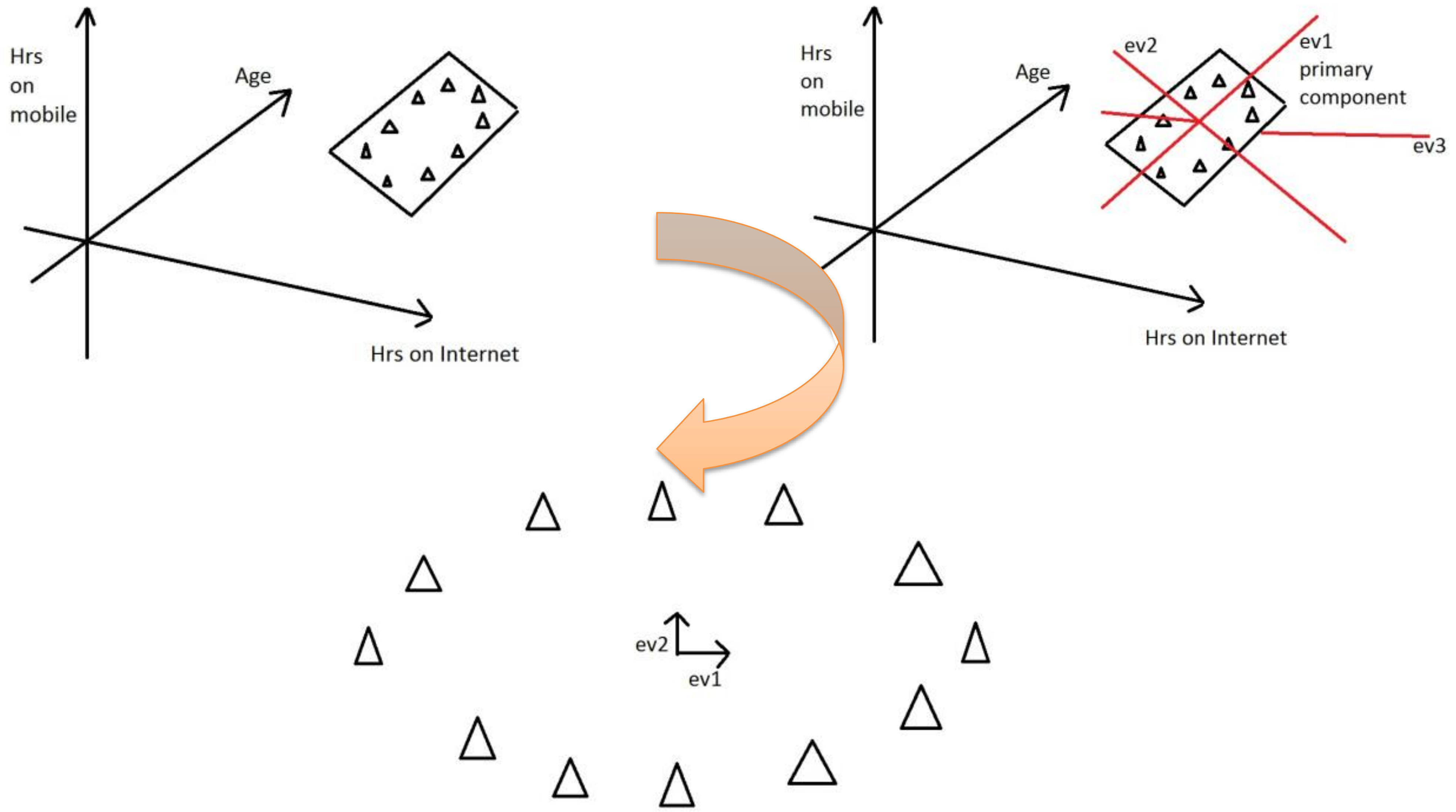


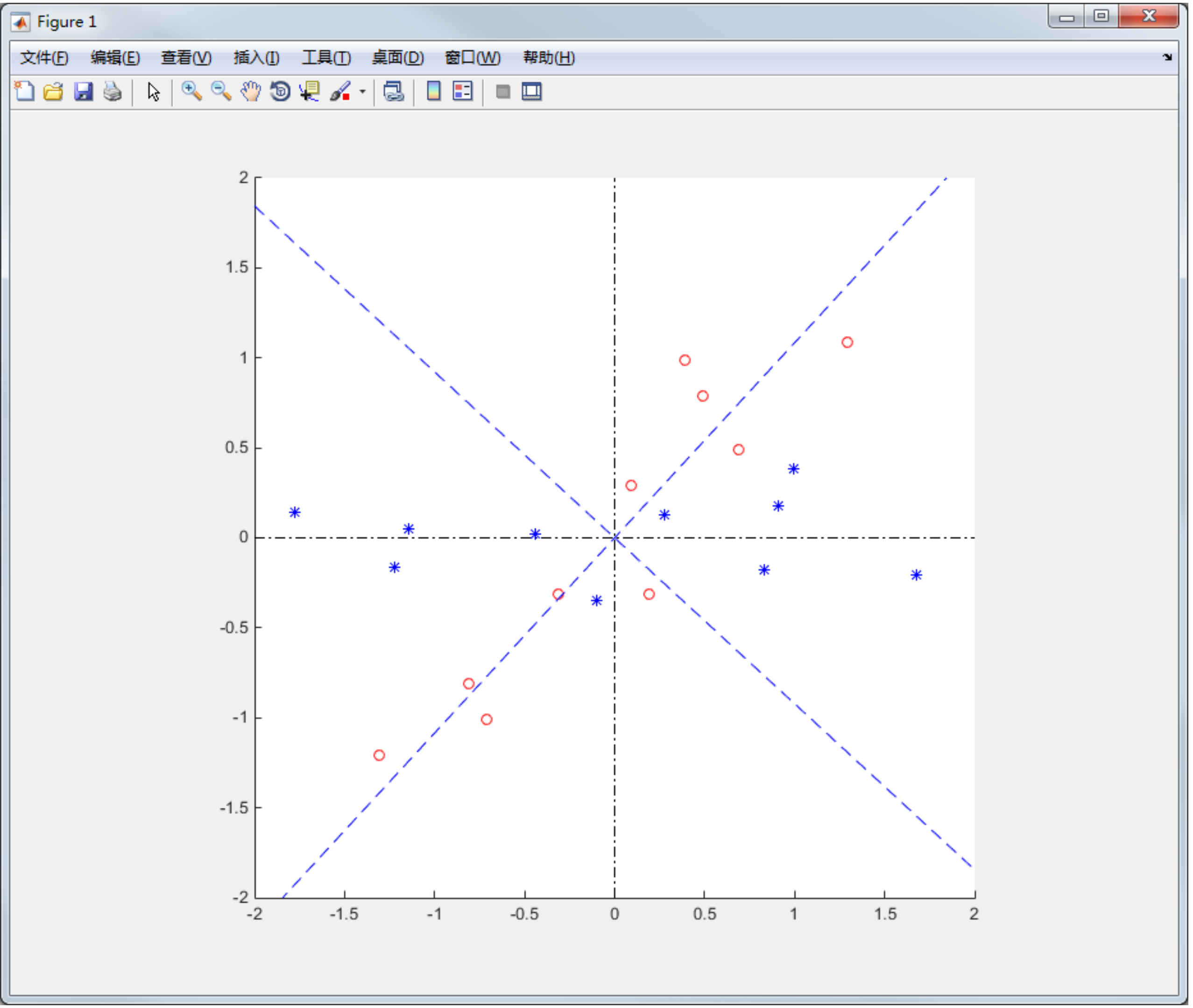


How to use PCA?



Dimension Reduction



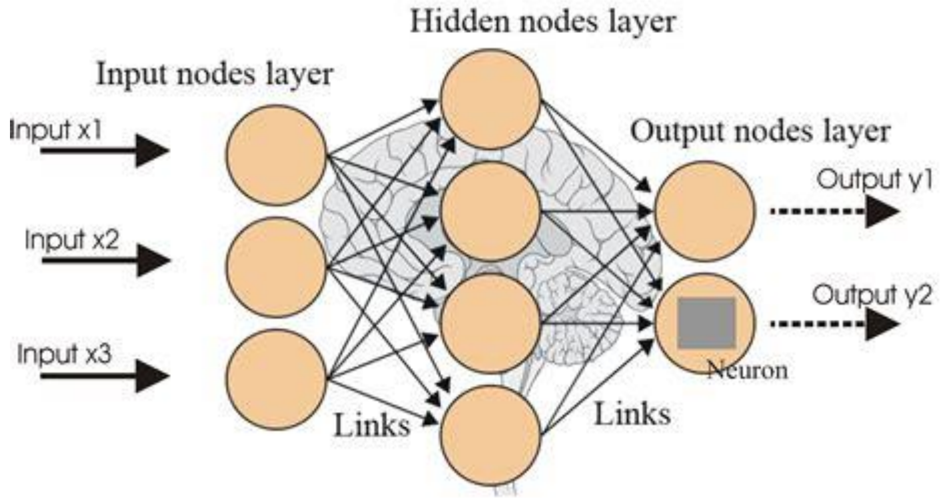


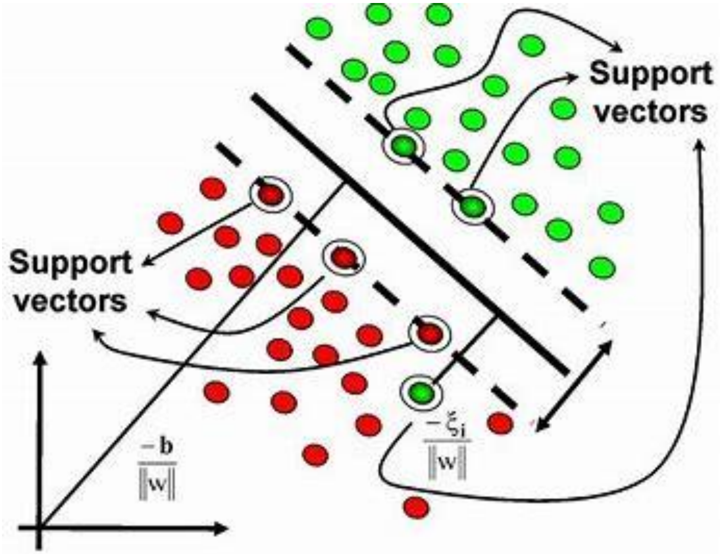

```
mat=[0.690000000000000,0.490000000000000;-1.31000000000000,-1.21000000000000;0.390000000000000,0.990000000000000;0.0900000000000000,0.290000000000000;1.29000000000000,1.09000000000000;0.490000000000000,0.790000000000000;0.190000000000000,-0.310000000000000;-0.810000000000000,-0.810000000000000;-0.310000000000000,-0.310000000000000;-0.710000000000000,-1.01000000000000];
hold on
axis square;
scatter(mat(:,1),mat(:,2),'o','r');
plot([-2 2],[0 0],'-.k');
plot([0 0],[-2 2],'-.k');
cov_mat=cov(mat);
[egvec,egval]=eig(cov_mat);
slope1 = egvec(2,1)/egvec(1,1);
slope2 = egvec(2,2)/egvec(2,1);
x=[-2:0.1:2];
vec1 = x*slope1;
vec2 = x*slope2;
xlim([-2,2]);
ylim([-2,2]);
plot(x,vec1,'--b');
plot(x,vec2,'--b');

egvec_sort = fliplr(egvec);
newmat = egvec_sort' * mat';
newmat =newmat';
scatter(newmat(:,1),newmat(:,2),'*','b');
```

SUSTech | Environment of Science

Supervised Classification

- Some advanced techniques
  - Neural networks
  - Support Vector Machines
  - Deep Learning
  - Contextual classifiers
    - Incorporate spatial or temporal conditions





Unsupervised Classification

- Recall: In unsupervised classification, the spectral data imposes constraints on our interpretation
- How? Rather than defining training sets and carving out pieces of n-dimensional space, we define no classes beforehand and instead use statistical approaches to divide the n-dimensional space into clusters with the best separation
- After the fact, we assign class names to those clusters