Technology Transition Workshop| *Mohamed R. Mahfouz, Ph.D.*

History and Application of Empirical Modeling

Outline

- **Empirical Models Vs. Theoretical Models**
- **Machine Learning**
  - **History**
  - **Applications**
  - **Types**
- **Regression Models**
  - **Parametric**
  - **Non Parametric**

Technology
Transition Workshop

Empirical Vs. Theoretical

- **Empirical statistics: calculating probability or information about an event from past experiment data**
  - Determined by data from an actual experiment
- **Theoretical statistics: calculating probability or information about an event based on sample space of known equally likely outcomes**
  - Determined by finding all the possible outcomes theoretically, and calculating how likely the given outcome is

Technology
Transition Workshop

NIJ National Institute of Justice

Machine Learning

- **What is Machine Learning?**

  "Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time"

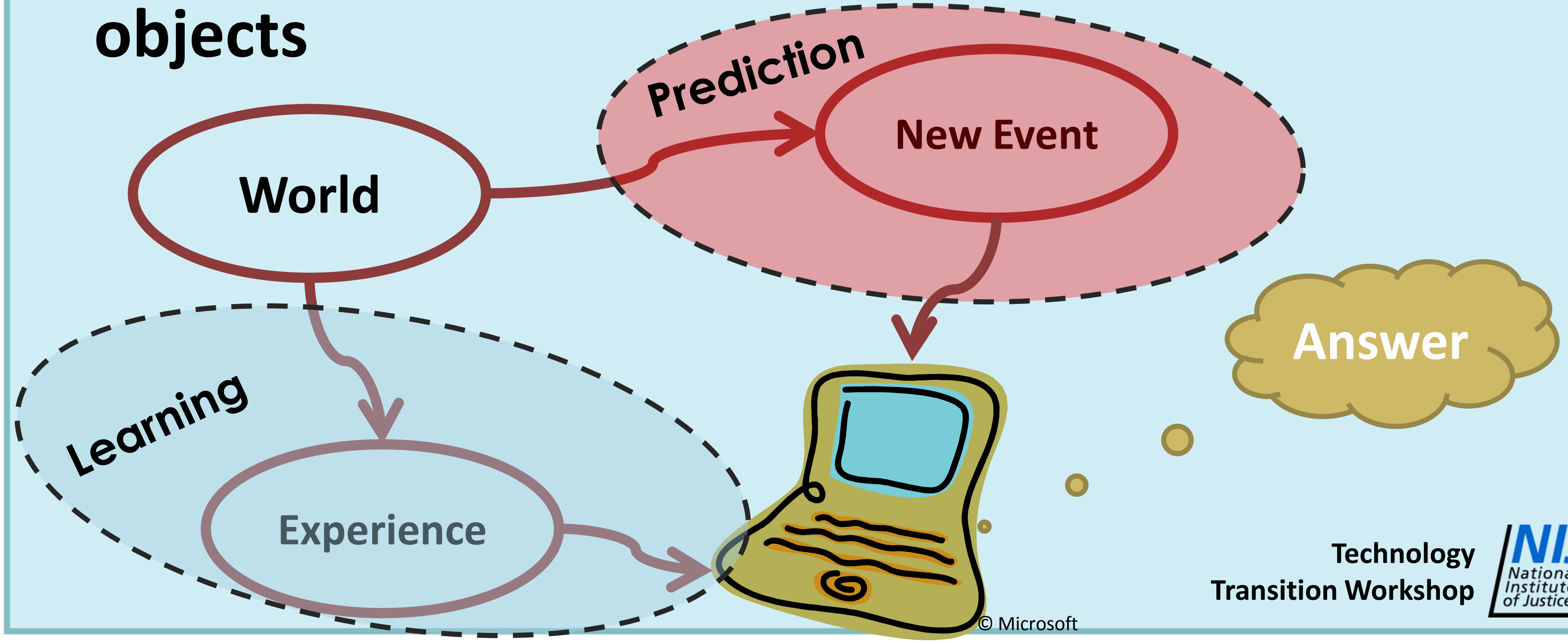
From Simon, H.A. (1983).

Technology Transition Workshop

NIJ National Institute of Justice

Machine Learning

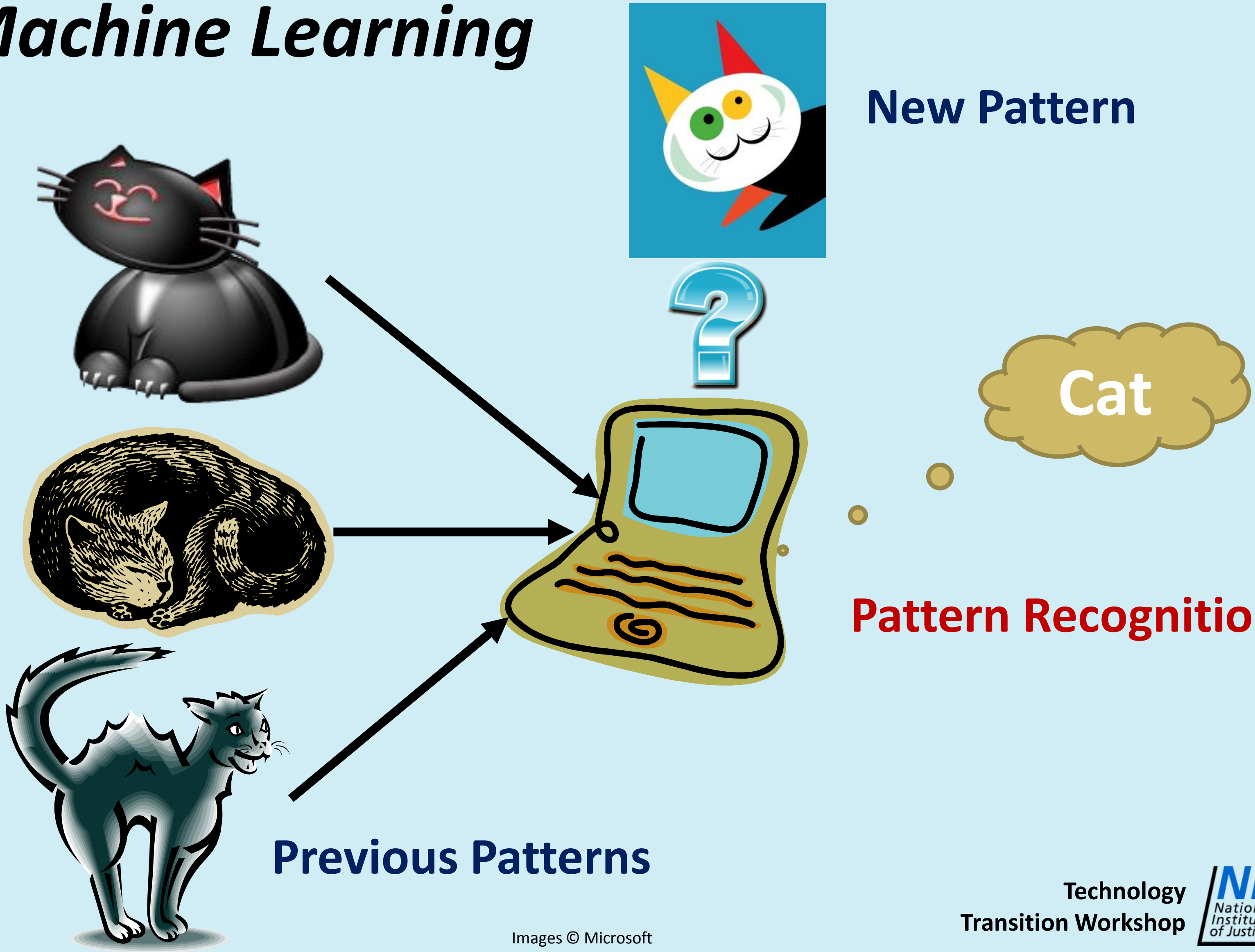
- **Machine learning is always driven by empirical data from the real world**
- **Goal is to teach computer how to use this data to predict future data or recognize certain events or objects**



Technology
Transition Workshop

NIJ National Institute of Justice

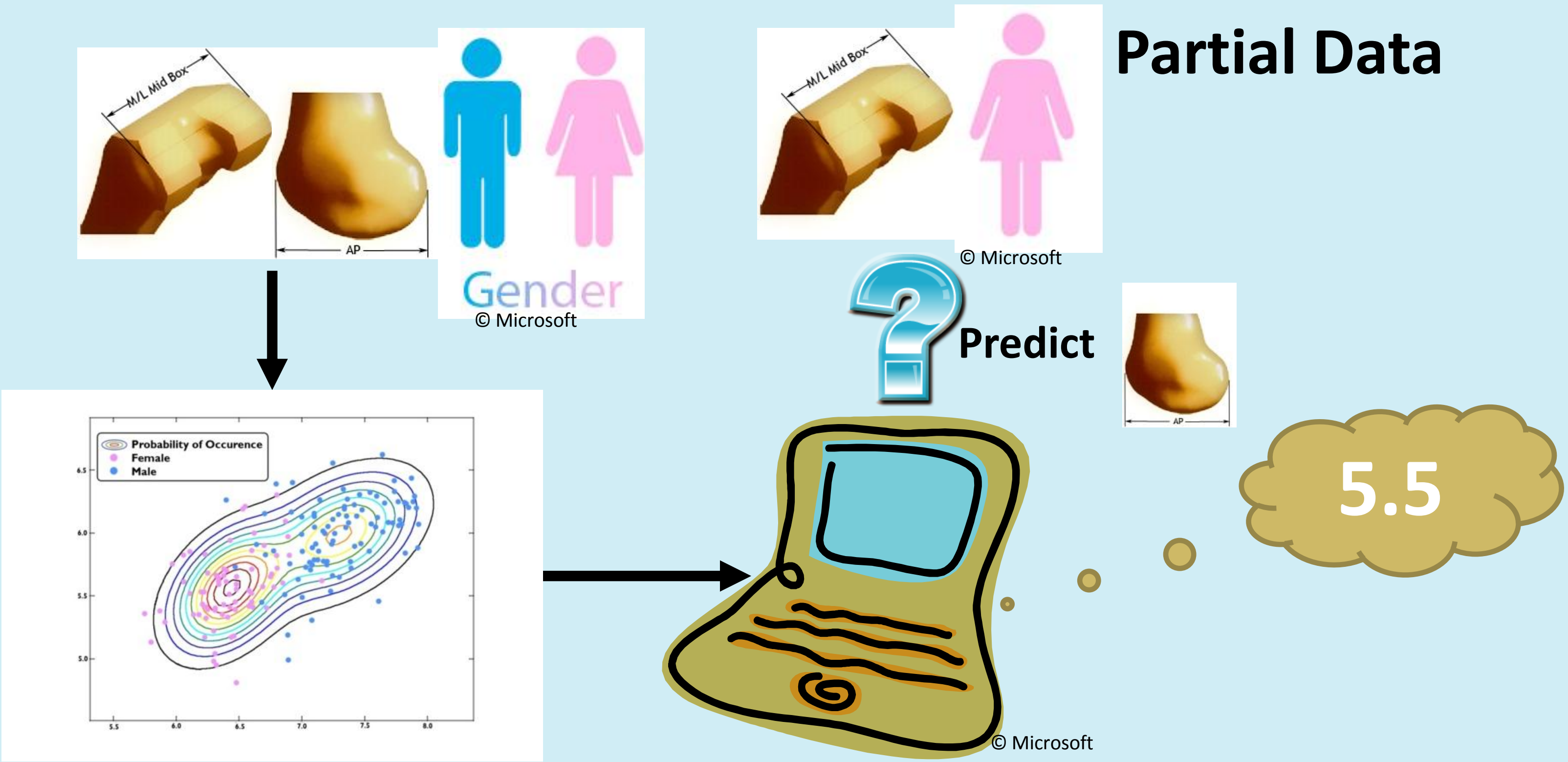
Machine Learning



Images © Microsoft

Technology
Transition Workshop
NIJ National Institute of Justice

Machine Learning



Experimental Data

Prediction Model

Machine Learning

- **Optimize a performance criterion using example data or past experience**

- **Role of statistics: inference from a sample**
- **Role of computer science: efficient algorithms to:**
  - **Solve the optimization problem**
  - **Represent and evaluate the model for inference**

Technology Transition Workshop

NIJ National Institute of Justice

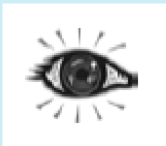
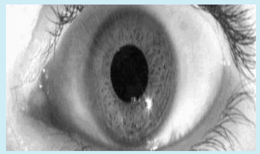
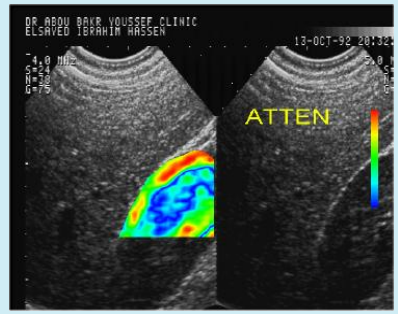
Machine Learning

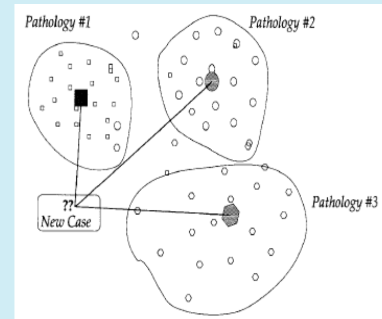
- **Human expertise is absent (navigating on Mars)**
- **Humans are unable to explain their expertise (speech recognition, vision, language)**
- **Solution changes in time (Stock market)**
- **Solution needs to be adapted to particular cases (user biometrics)**
- **The problem size is to vast for our reasoning capabilities (calculating webpage ranks)**

Technology Transition Workshop

NIJ National Institute of Justice

Machine Learning Applications

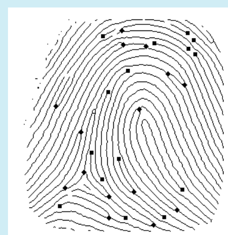
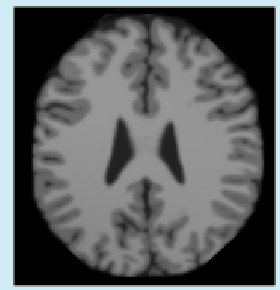
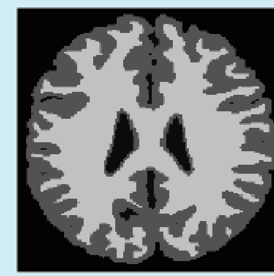






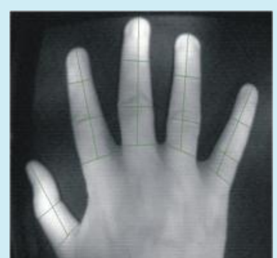
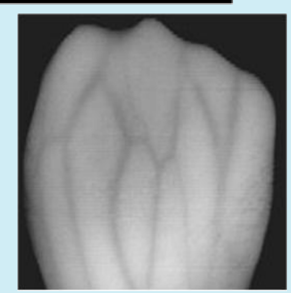
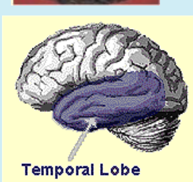


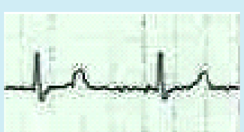
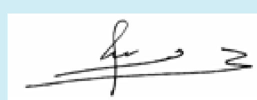






© Microsoft

Technology Transition Workshop

NIJ National Institute of Justice

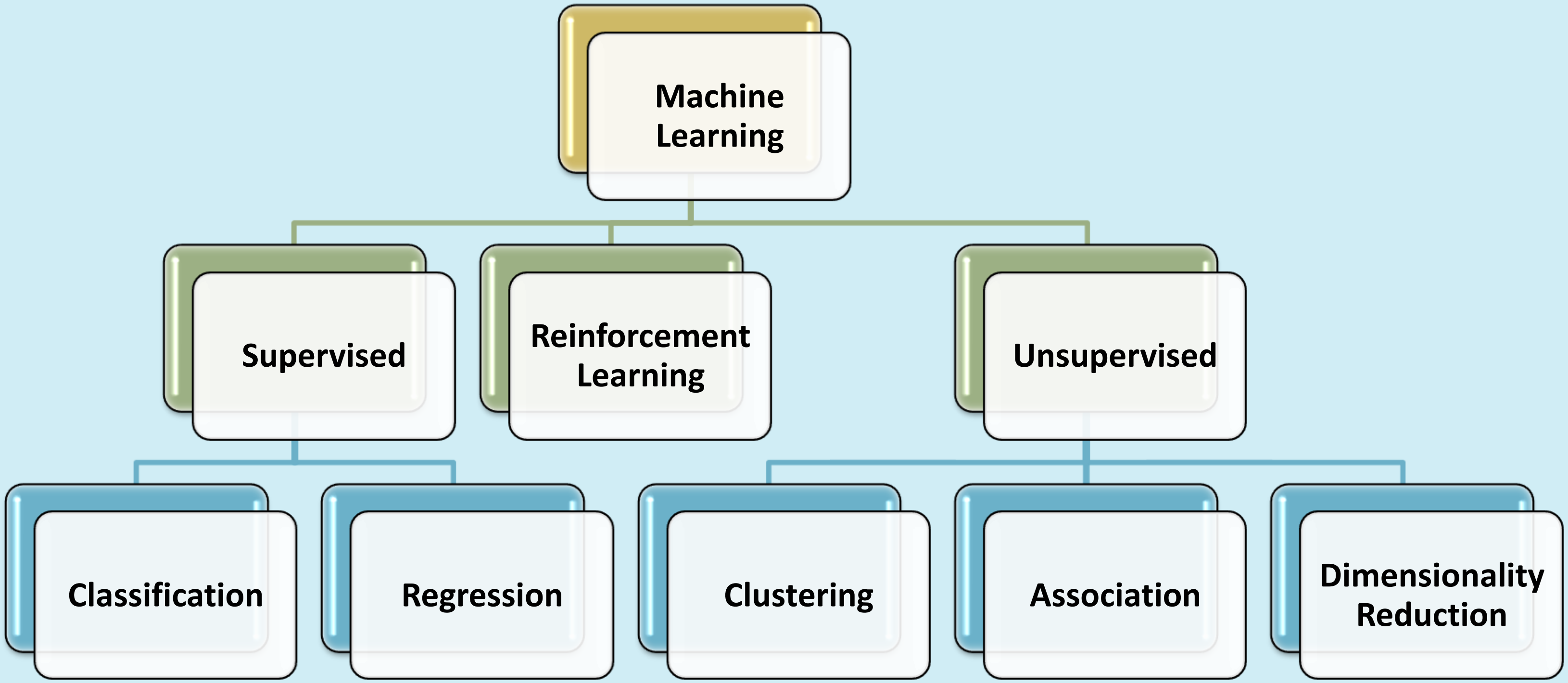
Machine Learning Applications

- **Retail:** Market basket analysis, customer relationship management
- **Biometrics:** Voice recognition, fingerprint, iris
- **Finance:** Credit scoring, fraud detection
- **Manufacturing:** Optimization, troubleshooting
- **Medicine:** Medical diagnosis
- **Telecommunications:** Optimizing service quality
- **Bioinformatics:** Sequencing
- **Web mining:** Search engines

Technology Transition Workshop

NIJ National Institute of Justice

Types of Learning

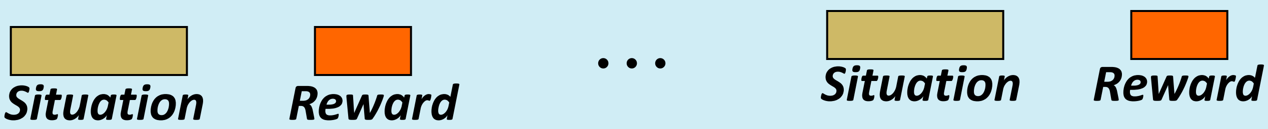


Technology
Transition Workshop

NIJ National Institute of Justice

Reinforcement Learning

- The agent acts on its environment, it receives some evaluation of its action (reinforcement), but is not told of which action is the correct one to achieve its goal
- Evaluation feedback can be seen as delayed reward for the system



- Example
  - Chess Game:
    - Reward is winning game at the end
  - Tennis Game:
    - Reward on each point scored
  - Dog Training:
    - Treat with every good behavior

Technology Transition Workshop
National Institute of Justice

Unsupervised Learning

- **No labels or feedback (No Expert)**
- **Studies how input patterns can be represented to reflect the statistical structure of the overall collection of input patterns**
- **No outputs are used (unlike supervised learning and reinforcement learning)**
  - **Clustering**
  - **Density estimation**

Technology Transition Workshop

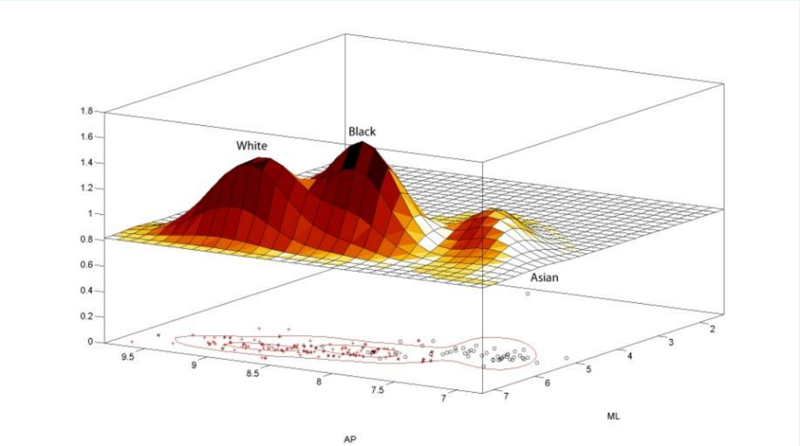
NIJ National Institute of Justice

Clustering

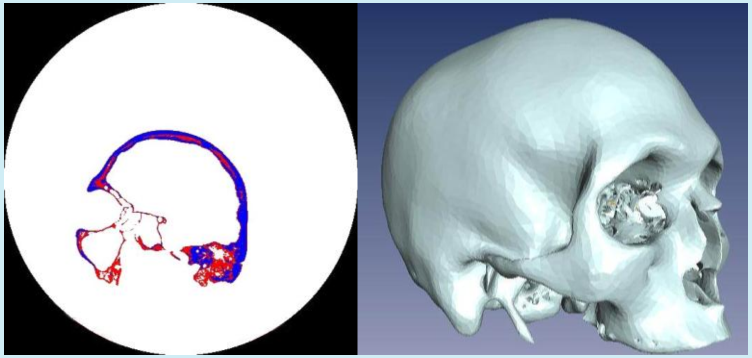
- **Types:**
  - **Hierarchical**
  - **Hard clustering**
    - **K-means**
  - **Soft clustering**
    - **Fuzzy c-means**

Technology
Transition Workshop

Clustering Applications



Optimum Number of Implant Sizes (k-means)



Segmentation (Fuzzy c-means)

Technology Transition Workshop

NIJ National Institute of Justice

Supervised Learning

- **Needs an expert**
- **Divided into two phases**
  - **Training:**
    - **For dataset A, expert provides labels f(x)**
    - **Goal – to find most probable model to generalize the training data**
  - **Testing:**
    - **Using inferred model, predict label of new point**

Technology
Transition Workshop

Examples of Supervised Learning

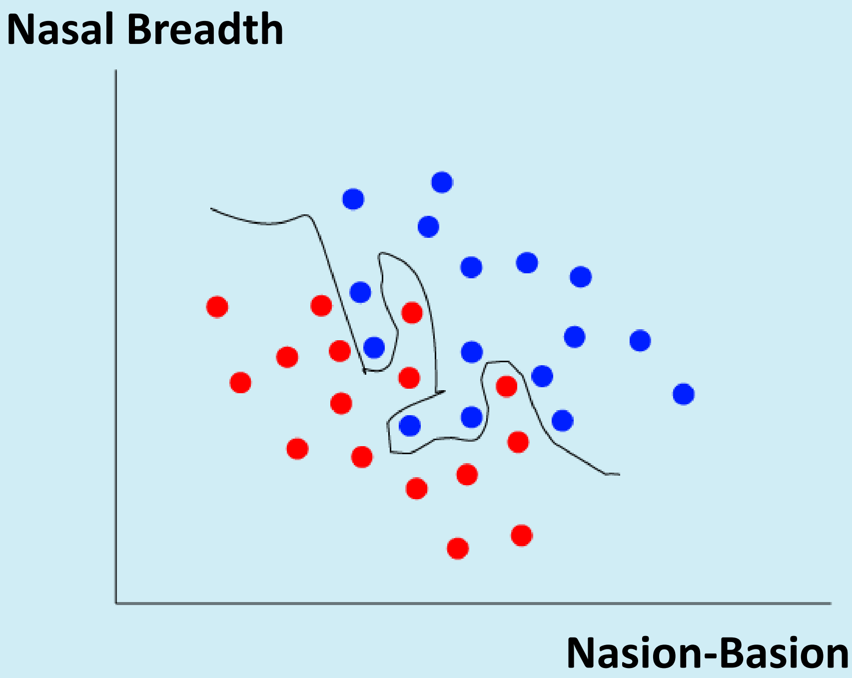
- **Handwriting recognition**
  - x: Data from pen motion
  - *f*(x): Letter of the alphabet
- **Character recognition**
  - x: Scanned document as image
  - f(x): Words in the document
- **Disease diagnosis**
  - x: Properties of patient (symptoms, lab tests)
  - f(x): Disease (or maybe recommended therapy)
- **Face recognition**
  - x: Picture of person's face
  - f(x): Person’s name
- **Spam detection**
  - x: Email message
  - f(x): Spam or not spam

Technology Transition Workshop

NIJ National Institute of Justice

Classification

- Example: Ancestry detection
- Differentiating between black and white skulls from two measurements



Technology Transition Workshop

NIJ National Institute of Justice

Classification Techniques

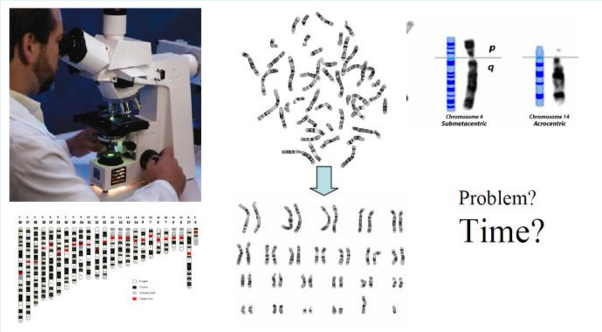
- **Linear**
  - **Perceptron learning**
  - **Linear discriminate**
  - **Support vector machines**
- **Nonlinear**
  - **Back-propagation neural network**
  - **Radial basis functions**
  - **Nonlinear support vector machine**
  - **Decision trees**

Technology
Transition Workshop

NIJ National Institute of Justice

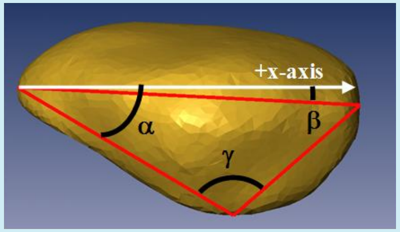
Classification Application

Chromosome Classification 97%



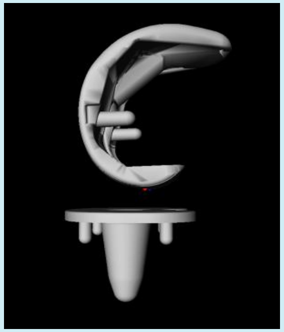
From Badawi, Hasan, Abdel Fatah, and Tadross (2003).

Patella Sexing 93.7 %



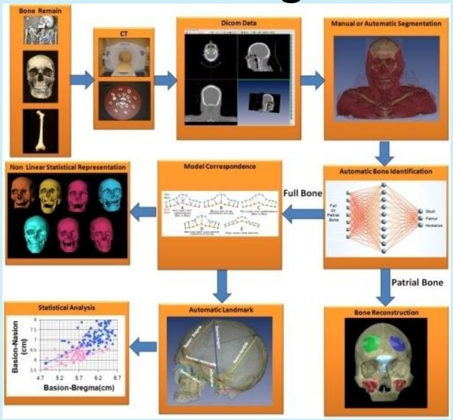
From Mahfouz, Badawi, Merkl, Abdel Fatah, Pritchard, Kesler, Moore, Jantz and Jantz (2007).

Kinematic Classification 90%



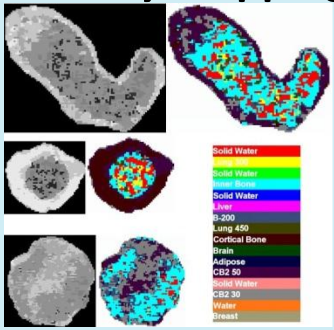
From Mahfouz, Battaglia, Abdel Fatah, Zingde, and Komistek (2011).

Skull Sexing 98%



From Shirley, Abdel Fatah, Jantz and Mahfouz (2011).

Density Mapping



From Moore, Mahfouz, Abdel Fatah and Badawi (2006).

Image Enhancement

From Abdel Fatah, Badawi, and Mahfouz(2006).

Technology
Transition Workshop

NIJ National Institute of Justice

Regression

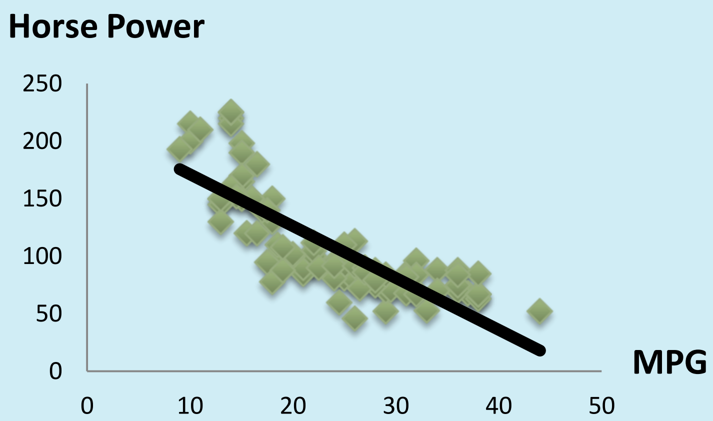
- Find functional description of data with the goal of predicting values for new input.
- Given data (x , y) = 1….n, find prediction function f(x) that can predict value of y from x.

$$\boldsymbol{f(x) = y}$$

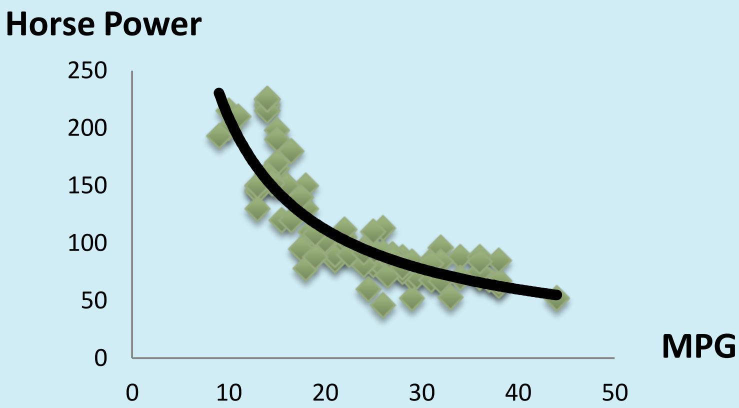
Technology
Transition Workshop

NIJ National Institute of Justice

Regression (Continued)



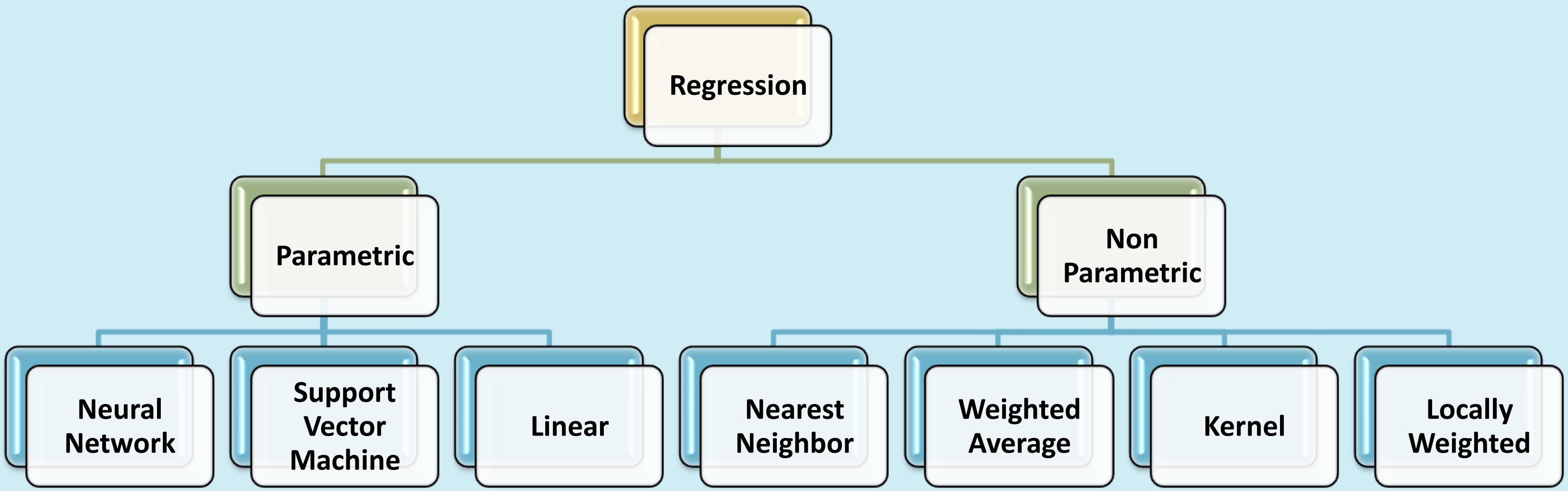
Linear Regression



Nonlinear Regression

Technology
Transition Workshop

Regression (Continued)



Technology Transition Workshop

Regression (Continued)

- **Parametric:**
  - **Includes linear regression, neural networks, support vector machine and other techniques that map relationships in data by optimizing different parameter values using a data set similar, but not exact to, future data sets**
  - **Once the parameters for the model are identified, the training data are no longer used and the model's prediction equation is set**
  - **Problems:**
    - **Over fitting**
    - **In the case of new data, model has to be retrained**

Technology Transition Workshop

NIJ National Institute of Justice

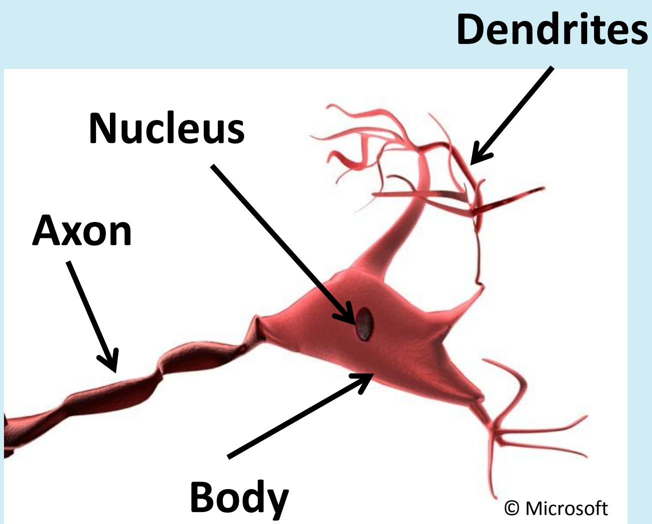
Neural Network

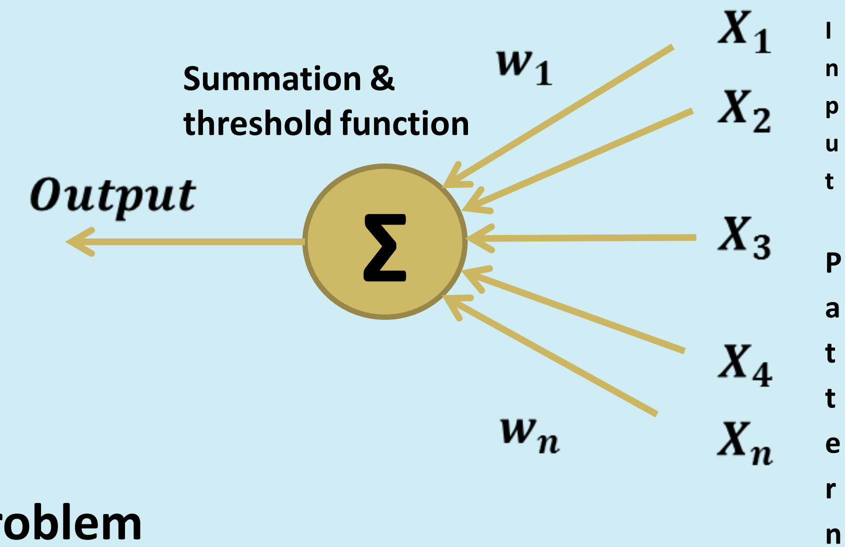
- Single Perceptron
  - Linear
  - Solve AND / OR problem fails XOR
- Multilayer back-propagation
  - Nonlinear
  - Consists of multiple layers (Input, Hidden, Output)
  - Create multiple hyper planes to solve complex space

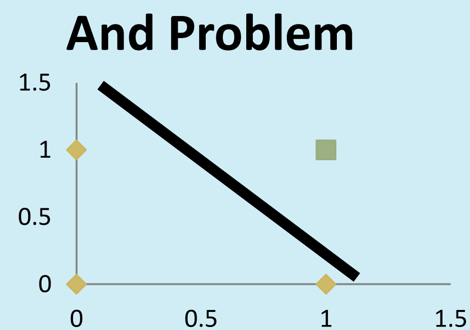
Technology
Transition Workshop

NIJ National Institute of Justice

Single Perceptron



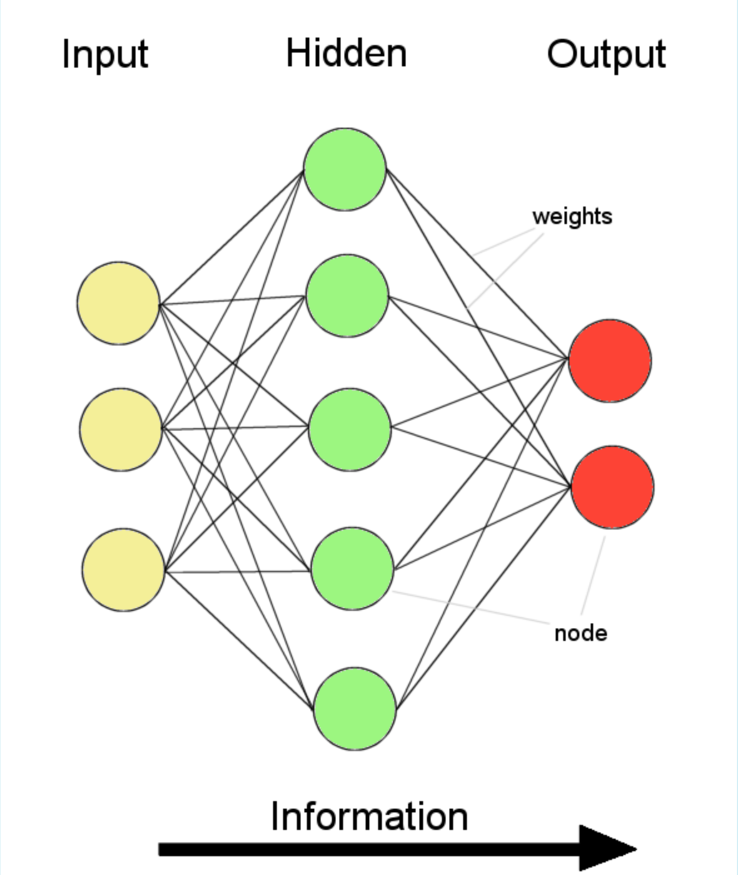


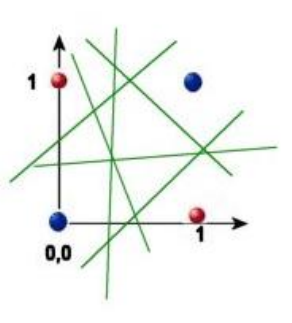


Technology
Transition Workshop

Multilayer Back Propagation

- **Information flow is unidirectional**
  - **Data is presented to Input layer**
  - **Passed on to Hidden Layer**
  - **Passed on to Output layer**
- **Information is distributed**
- **Information processing is parallel**
- **Requires retraining if new pattern presented**



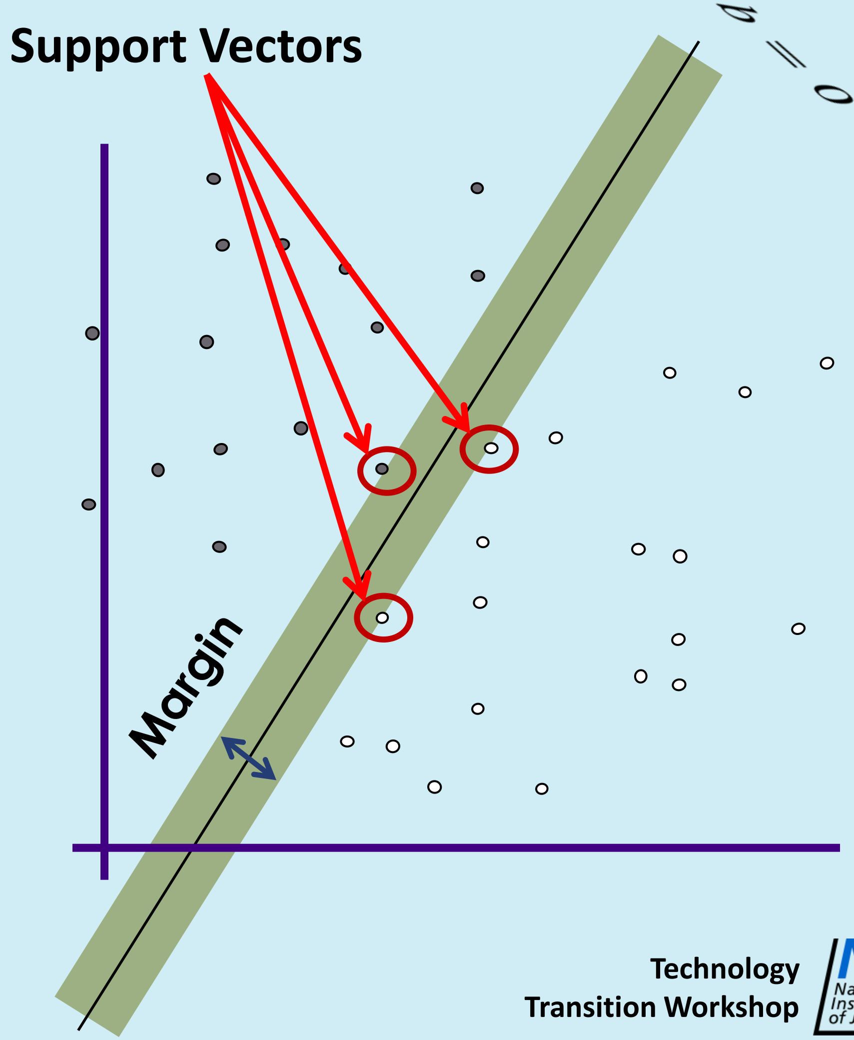


Technology Transition Workshop

NIJ National Institute of Justice

Support Vector Machine

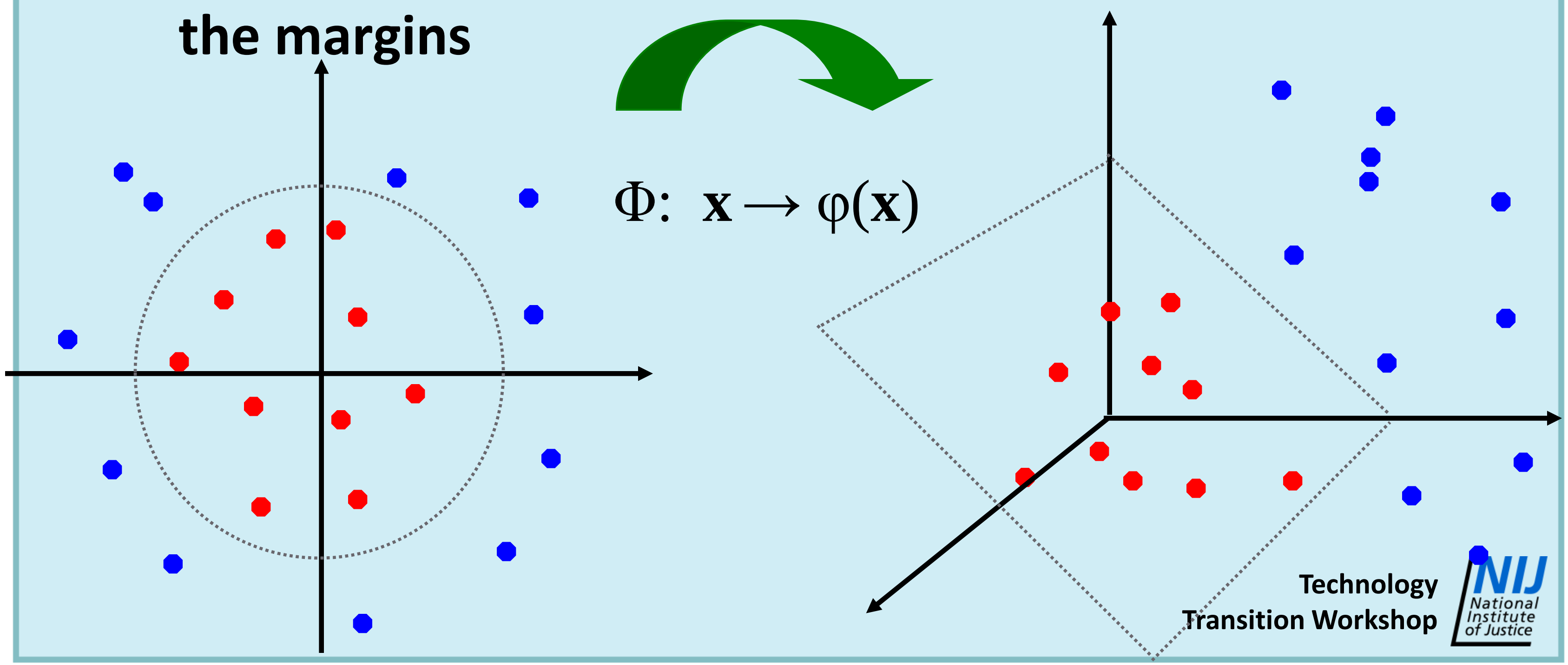
- Linear classifier
- Find plane that maximizes margin
- Support vectors are the points which the margin pushes against



Technology Transition Workshop
NIJ National Institute of Justice

Nonlinear Support Vector Machine

- **Using Kernel Trick**
  - **Maps data to higher dimension space where it is separable**
  - **Then apply SVM to find hyper planes that maximize the margins**



Technology Transition Workshop

NIJ National Institute of Justice

Non Parametric Regression

- Use the actual training data to understand future predictions and store the training data in a "memory matrix"
- Rather than modeling the whole input space with a parametric model such as a neural network or linear regression, local techniques construct a local model in the immediate region of the query
- These models are constructed "on the fly", not beforehand

Technology
Transition Workshop

NIJ National Institute of Justice

Non Parametric Regression (Continued)

- **When the query is made, the algorithm locates training input patterns in its vicinity and performs a weighted regression with the similar observations**
- **The observations are weighted with respect to their proximity to the query point**
- **In order to construct a robust local model, one must define a distance function to measure what is considered to be local to the query, implement locally weighted regression, and consider smoothing techniques such as regularization**

Technology Transition Workshop

NIJ National Institute of Justice

Non Parametric Regression (Continued)

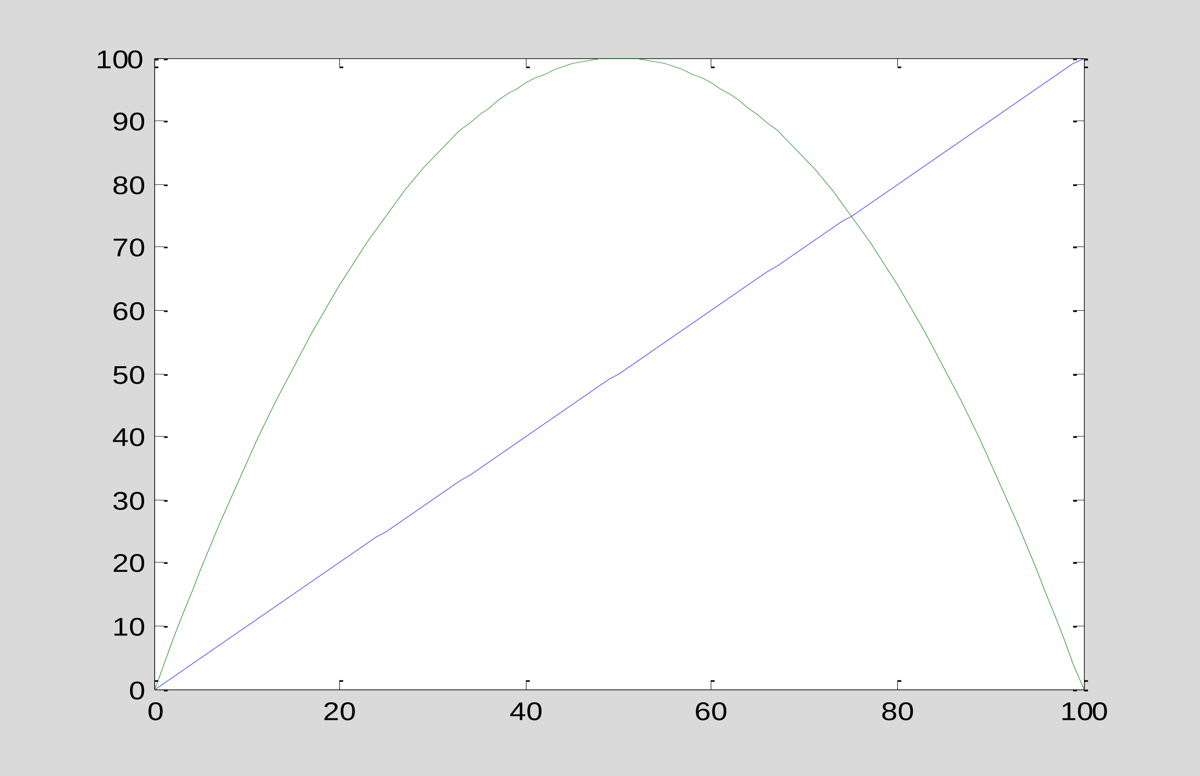
- **Nearest neighbor**
- **Weighted averaging**
- **Locally weighted regression**
- **Kernel regression**

Technology
Transition Workshop

Non Parametric Regression (Continued)

- Let's compare these methods using two example functions:

$$y_1 = x \qquad\qquad y_2 = 4x - \frac{1}{25}x^2$$



Technology
Transition Workshop

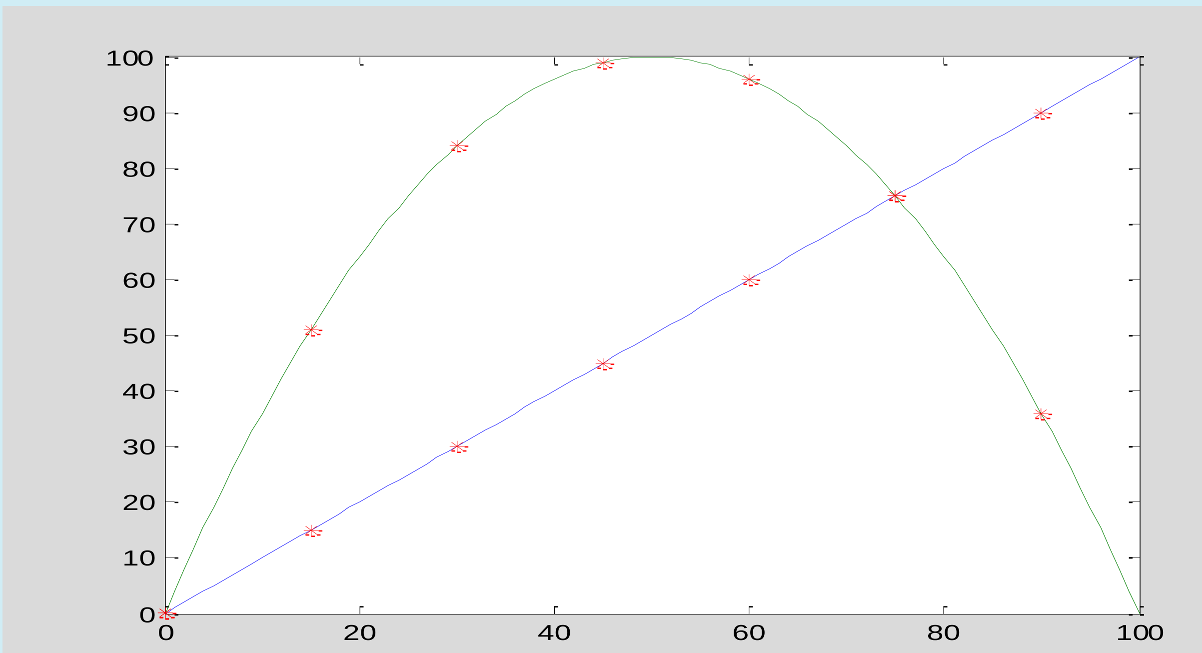
Non Parametric Regression (Continued)

- **We will develop a model using a few training points**

  xt=[0 15 30 45 60 75 90]'

- **and test the model with the complete data set**

  x=[0:1:100]



Technology Transition Workshop

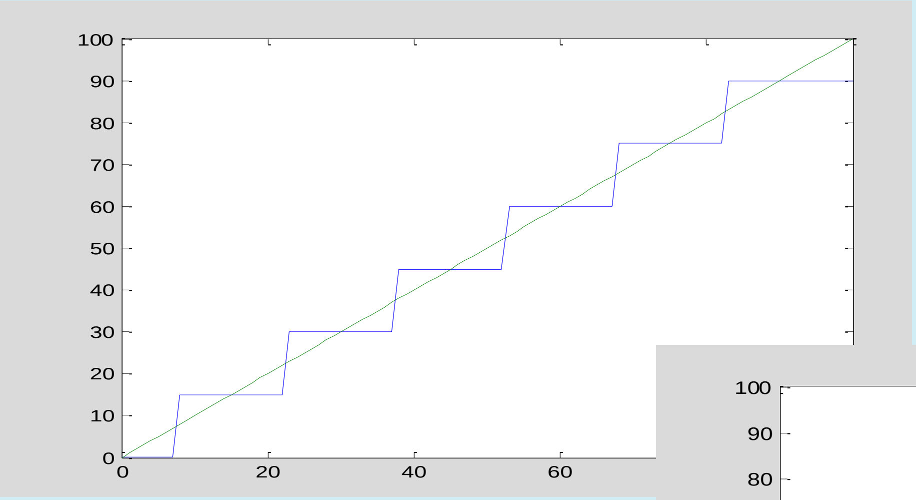
NIJ National Institute of Justice

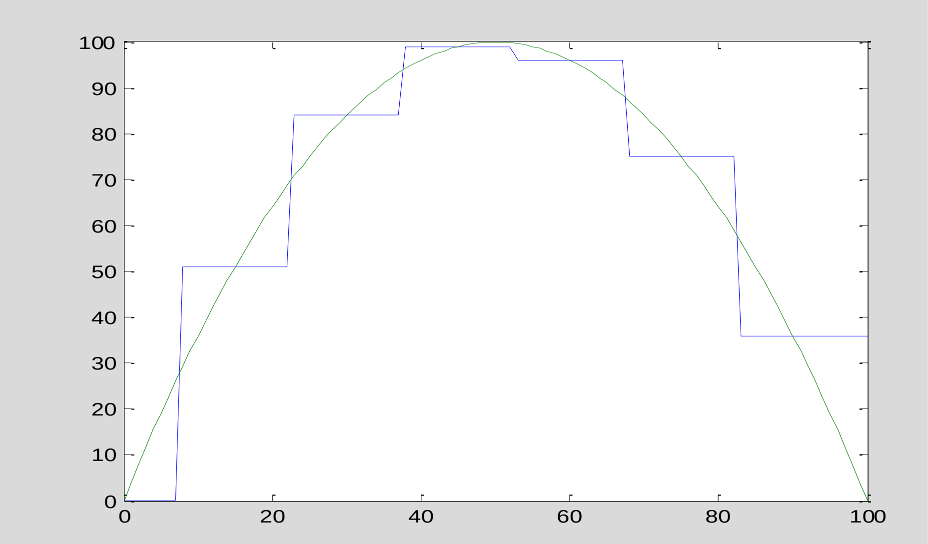
Nearest Neighbor

- If we were to predict the true value of a potentially noisy data point, we may want to estimate it with the nearest neighbor of a saved data set
- The nearest neighbors can be calculated using some distance measure, such as a Euclidian distance
- We will compare the functions using the nearest neighbor
- This is a very crude method

Technology
Transition Workshop

Nearest Neighbor (Continued)





Technology
Transition Workshop

Weighted Averaging

- **Now let's look at a weighted average where the output of nearby points are inversely weighted with respect to their distance from the query point (q):**

$$\hat{y}(q) = \frac{\sum_{i=1:n} \left(y_i * w_i \right)}{\sum_{i=1:n} w_i} \qquad\qquad w_i = \frac{1}{d(x,q)}$$

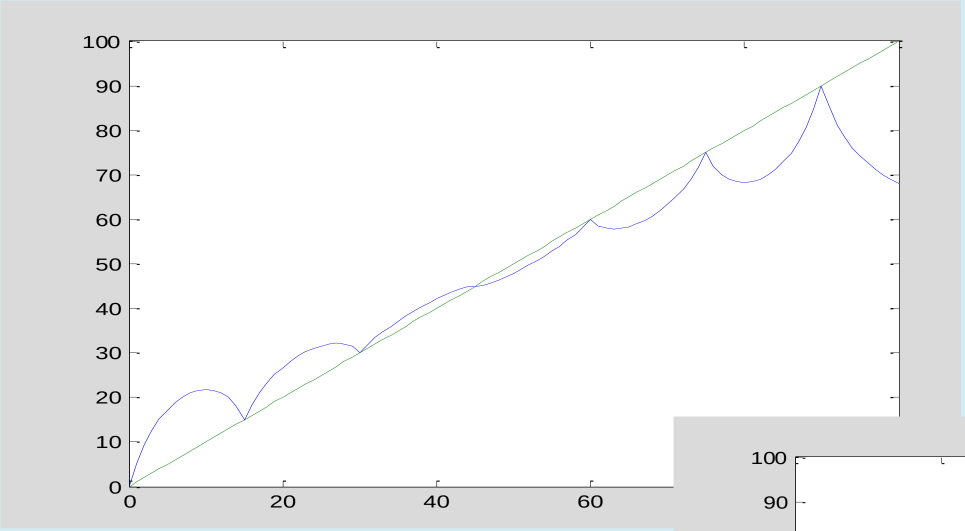
- **The most common distance function is the Euclidean distance:**

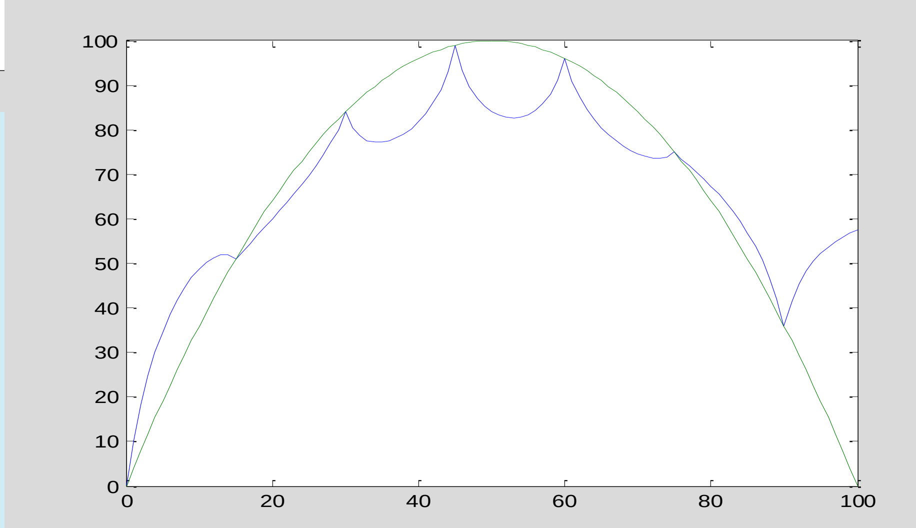
$$d(x,q) = \sqrt{\sum_j \left(x_j - q_j \right)^2}$$

Technology
Transition Workshop

NIJ National Institute of Justice

Inverse Weighted Average





Technology Transition Workshop

NIJ National Institute of Justice

Locally Weighted Regression

- Linear regression solves the following linear model:

$$y = X\beta + \varepsilon$$

- Where:
  - *y* is a vector (nx1) of samples of the response variable
  - *X* is a matrix (nxp) of predictor variables where the columns are the variables
  - the rows are the samples or observations
  - β is the vector of regression coefficients (px1) that linearly combines the predictors to form the response
  - ε is a vector (nx1) of the prediction errors
- Solve the weighted least squares regression equation for the optimal estimates of the regression coefficients

Technology
Transition Workshop

NIJ National Institute of Justice

Locally Weighted Regression (Continued)

- We can solve the weighted least squares regression equation for the optimal estimates of the regression coefficients:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{Y} = \left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{v}$$

- Where:

$$\mathbf{Z} = \mathbf{W}\mathbf{X}$$
$$and$$
$$\mathbf{v} = \mathbf{W}\mathbf{Y}$$

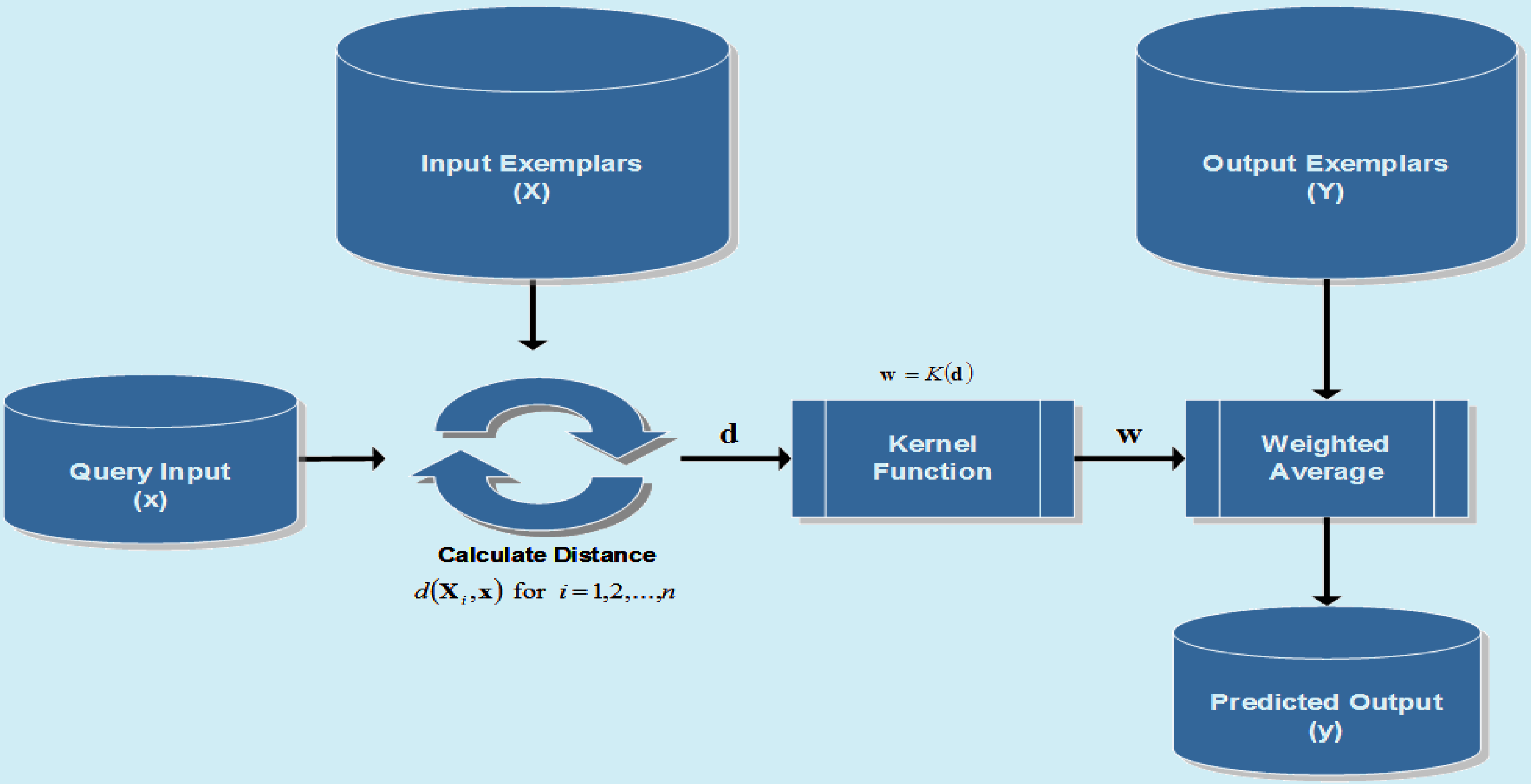
- With W being a diagonal matrix, with the diagonals being equal to the square root of the Kernel function:

$$w_{ii} = \sqrt{K(d(x_i, q))}$$

Technology
Transition Workshop

NIJ National Institute of Justice

Kernel Regression



Technology
Transition Workshop

Kernel Regression (Continued)

- **Kernel computes the regression coefficients as weights described by density function, with a scale parameter that adjusts the size and the form of the weights near q:**

$$\hat{y}(q) = \frac{\sum_{i=1:n} y_i w_i}{\sum_{i=1:n} w_i} = \frac{\sum_{i=1:n} y_i K(d(x_i, q))}{\sum_{i=1:n} K(d(x_i, q))}$$

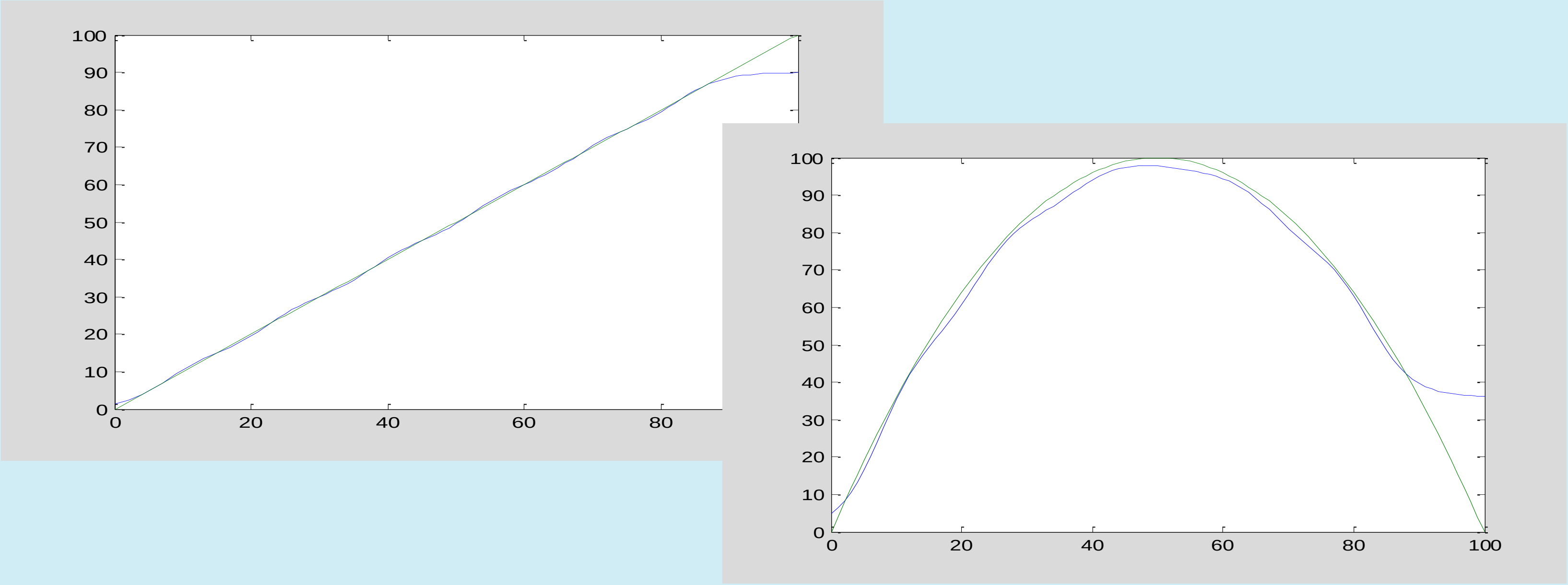
- **The most common Kernel function (K(d)) is the Gaussian kernel:**

$$w_i = K(d) = e^{-d^2}$$

Technology Transition Workshop

NIJ National Institute of Justice

Kernel Regression (Continued)

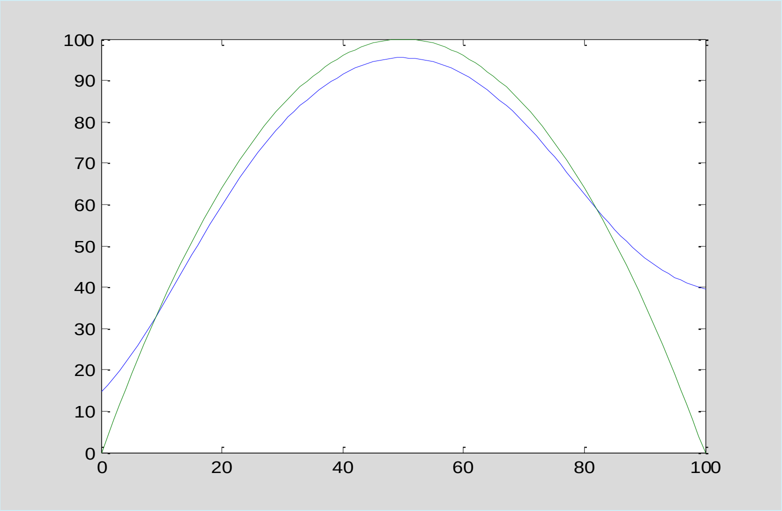


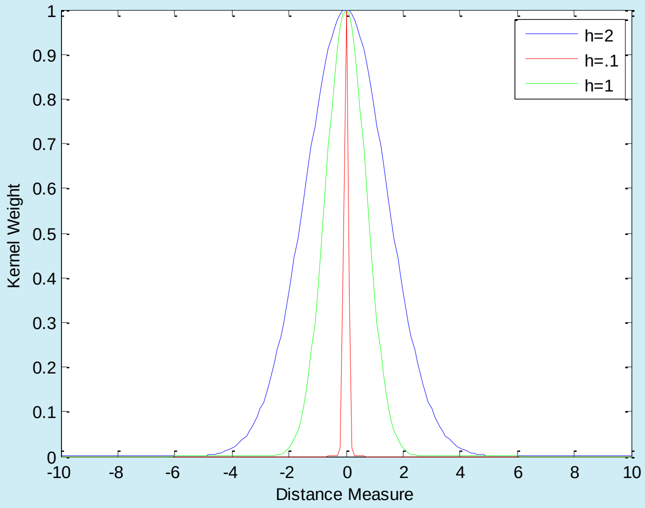
- **The smoothness of the function can be controlled with the bandwidth**
- **The bandwidth should always be chosen to be large enough to cover neighboring points**

Technology Transition Workshop

Kernel Regression (Continued)

- **A larger bandwidth acts as a regularization parameter by smoothing the solution**
- **It also introduces a bias**



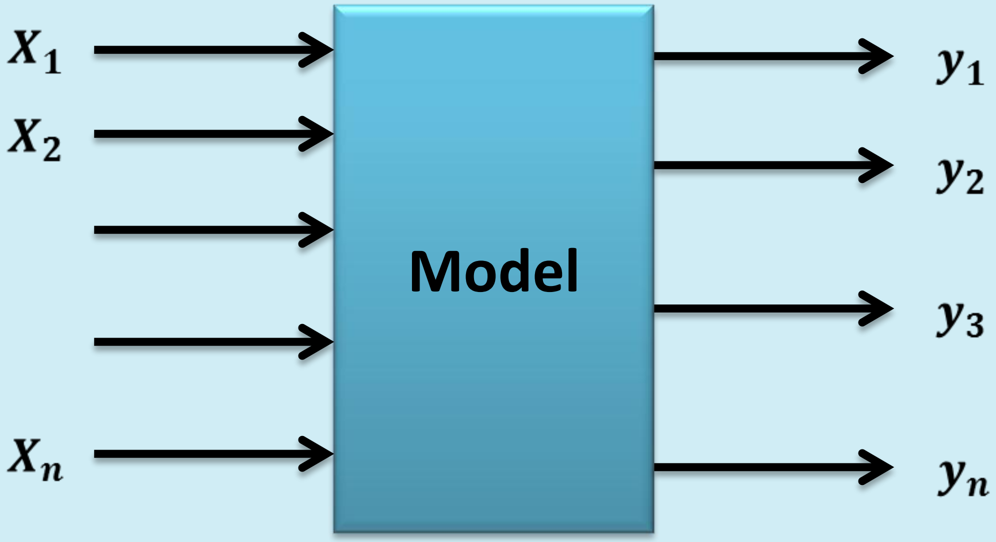


Technology Transition Workshop

NIJ National Institute of Justice

Kernel Regression *(Hetero Associative Model)*

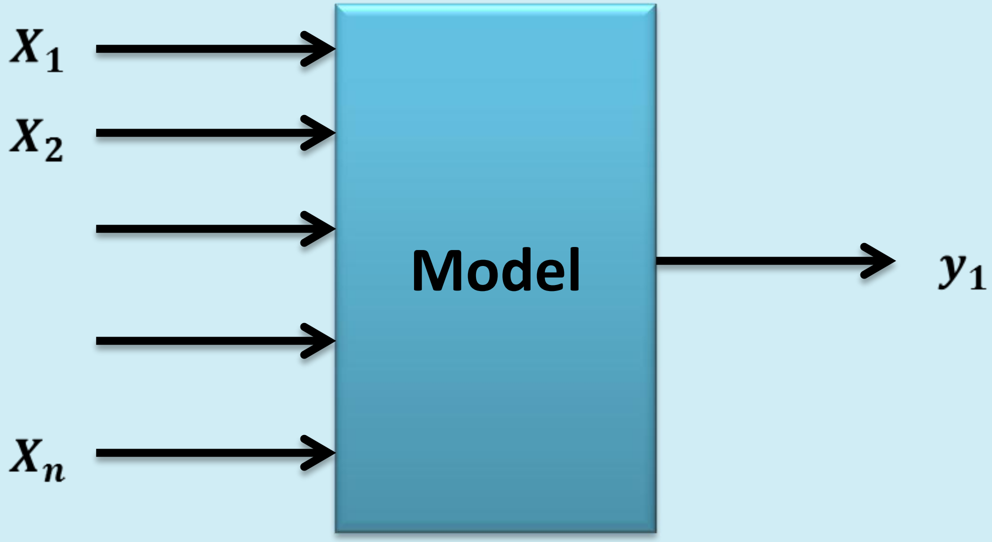
- **Number of inputs does not equal the number of outputs in the model**



Technology
Transition Workshop

Kernel Regression *(Inferential Model)*

- Number of outputs from the model is one
- Therefore, a separate KR model is constructed for each output exemplar



Technology
Transition Workshop

Cited Scientific References

- **Abdel Fatah, E.E.; Badawi, A.M.; Mahfouz, M.R. Fast 3D CT Diffusion Enhancement: A Neural Networks Filter. *Computer Methods in Biomechanics and Biomedical Engineering, 7th International Symposium*, Cote De Azure, France, 2006.**
- **Badawi, A.M.; Hasan, K.G.; Abdel Fatah, E.E.; Tadross, R. Chromosomes Classification Based on Neural Networks, Fuzzy Rule Based, and Template Classifiers. *In Proceedings of The 46th IEEE International Midwest Symposium of Circuits and Systems (MWSCAS)*, Cairo, Egypt, 2003.**
- **Mahfouz, M.R.; Badawi, A.M.; Merkl, B.C.; Abdel Fatah, E.E.; Pritchard, E.; Kesler, K.; Moore, M.K.; Jantz, R.L.; Jantz, L.M. Patella Sex Determination by 3D Statistical Shape Models and Nonlinear Classifiers. *Forensic Science International* 2007, *173(2)*, 161-170.**
- **Mahfouz, M.R.; Battaglia, N.; Abdel Fatah, E.E.; Zingde, S.M.; Komistek, R.D. Total Knee Replacement Implant Classification Using Neural Networks. *78th Annual Meeting of the American Academy of Orthopaedic Surgeons*, San Diego, CA, 2011.**

Technology Transition Workshop

NIJ National Institute of Justice

Cited Scientific References (Continued)

- **Moore, M.K.; Mahfouz, M.R.; Abdel Fatah, E.E.; Badawi, A.M. Comparison of Dry and In Vivo Bone Densities from CT Images using Fuzzy Logic and Neural Networks. *Computer Methods in Biomechanics and Biomedical Engineering, 7th International Symposium*, Cote De Azure, France, 2006.**
- **Shirley, N.R.; Abdel Fatah, E.E.; Jantz, R.L.; Mahfouz, M.R. Improving Sex Estimation from the Human Cranium Using 3D CT Scans. *American Academy of Forensic Sciences, 63rd Annual Scientific Meeting*, Chicago, IL, 2011.**
- **Simon, H.A. Why Should Machines Learn? *In Machine Learning: An Artificial Intelligence Approach*; Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. (Eds.). Tioga Publishing Co.: Palo Alto, CA, 1983; Chapter 2.**

Technology
Transition Workshop

NIJ National Institute of Justice

Questions?

Technology Transition Workshops are a project of NIJ's Forensic Technology Center of Excellence, operated by the National Forensic Science Technology Center (www.nfstc.org), funded through cooperative agreement #2010-DN-BX-K210.

These training materials are only for the course instructors and course participants and are for purposes associated solely for this course. Some of the materials may be subject to copyrights held by third parties. None of these materials may be: a) further disseminated or b) accessed by or made available to others. Individuals with questions concerning the permissibility of using these materials are advised to consult NFSTC at info@nfstc.org.

Technology
Transition Workshop

NIJ National Institute of Justice

Contact Information

Mohamed R. Mahfouz, Ph.D.

307 Perkins Hall

University of Tennessee

Knoxville, TN, 37996

865-974-7668

mmahfouz@utk.edu

Note: All images are courtesy of Dr. Mohamed R. Mahfouz and the Center for Musculoskeletal Research at The University of Tennessee, unless otherwise indicated.

Technology Transition Workshop

NIJ National Institute of Justice