



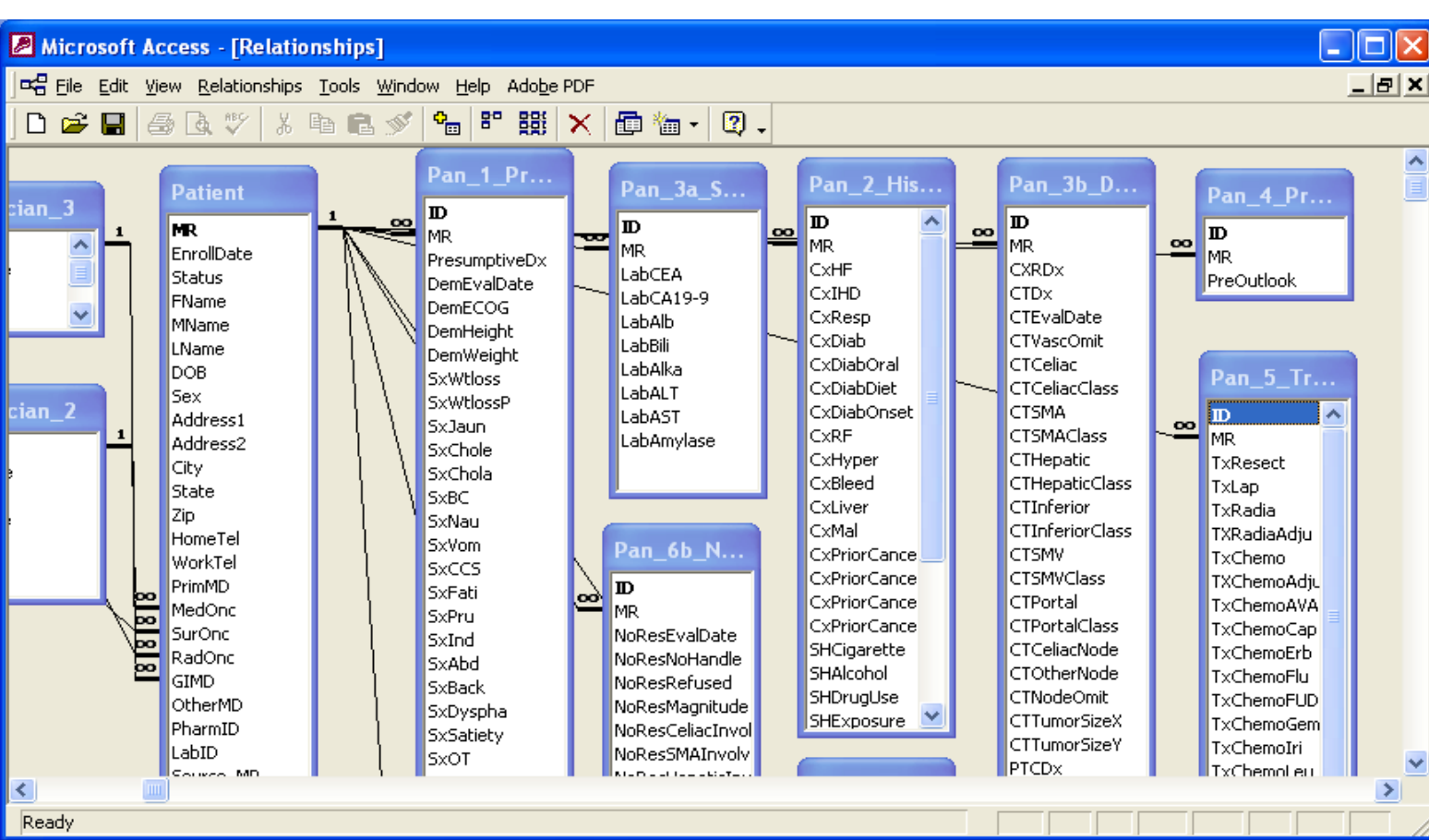
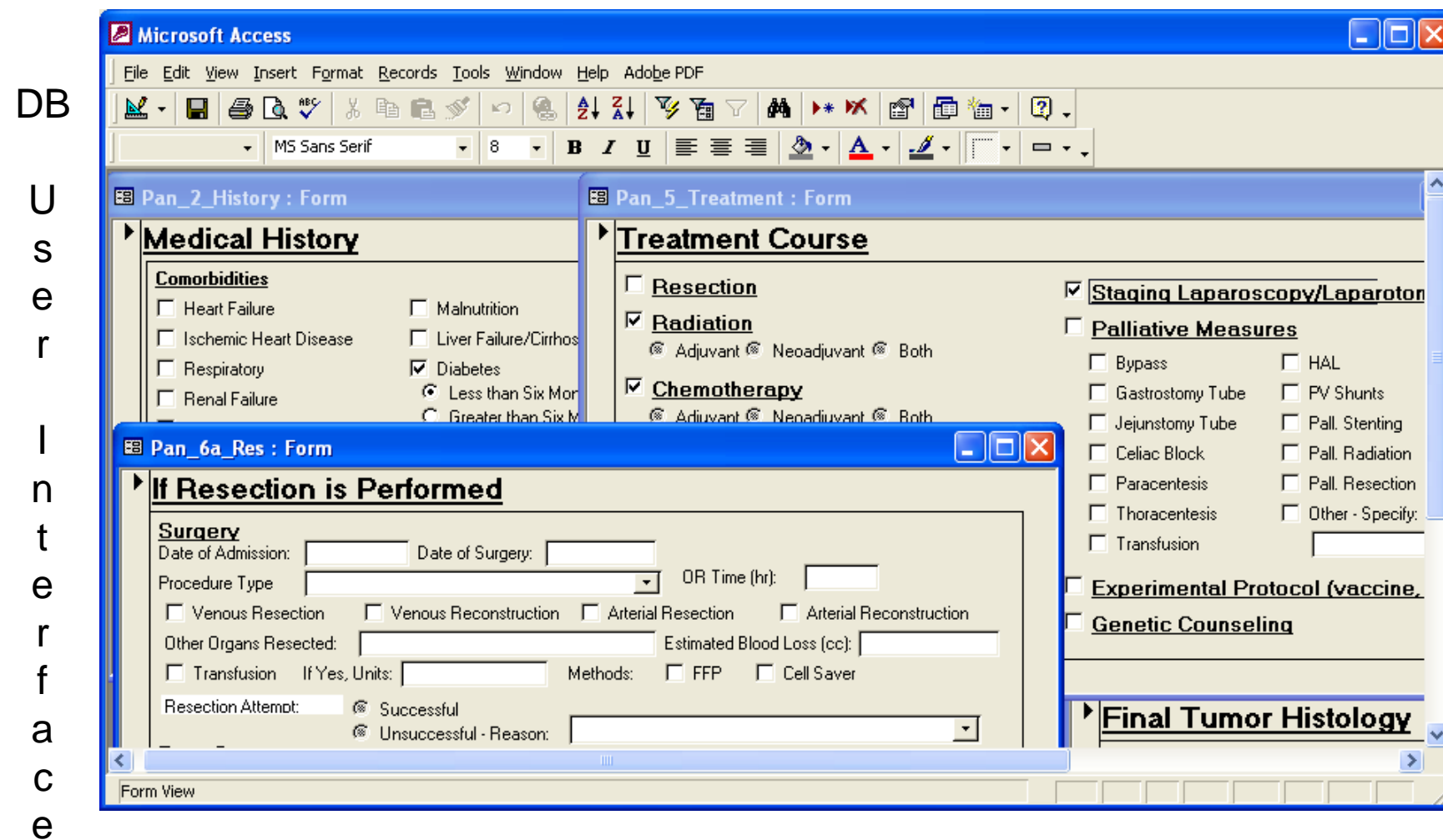
Machine Learning, Meta-Learning and Dimensionality Reduction for Pancreatic Cancer Patient Data

By Keith A. Pray (Computer Science) and Prof. Carolina Ruiz (Computer Science)

Goals: Improve current accuracy of predicting patient longevity and quality of life

1. Build upon previous pancreatic research work at WPI [Hayward 2006, Floyd 2007]

2. Apply data dimensionality reduction techniques



Database Captures Heterogeneous Structure of Patient Narrative

Currently about 200 data attributes for 248 patients

- Biographical
- Presentation of symptoms at time of diagnosis
- Medical history
- Diagnostic test results
- Medical imaging scans
- Preliminary evaluation
- Treatment details
- Surgical resection details or reasons not pursued
- Pathology report
- Follow up visit

Prediction Goals:

- Survival time from time of diagnosis
- Quality of life

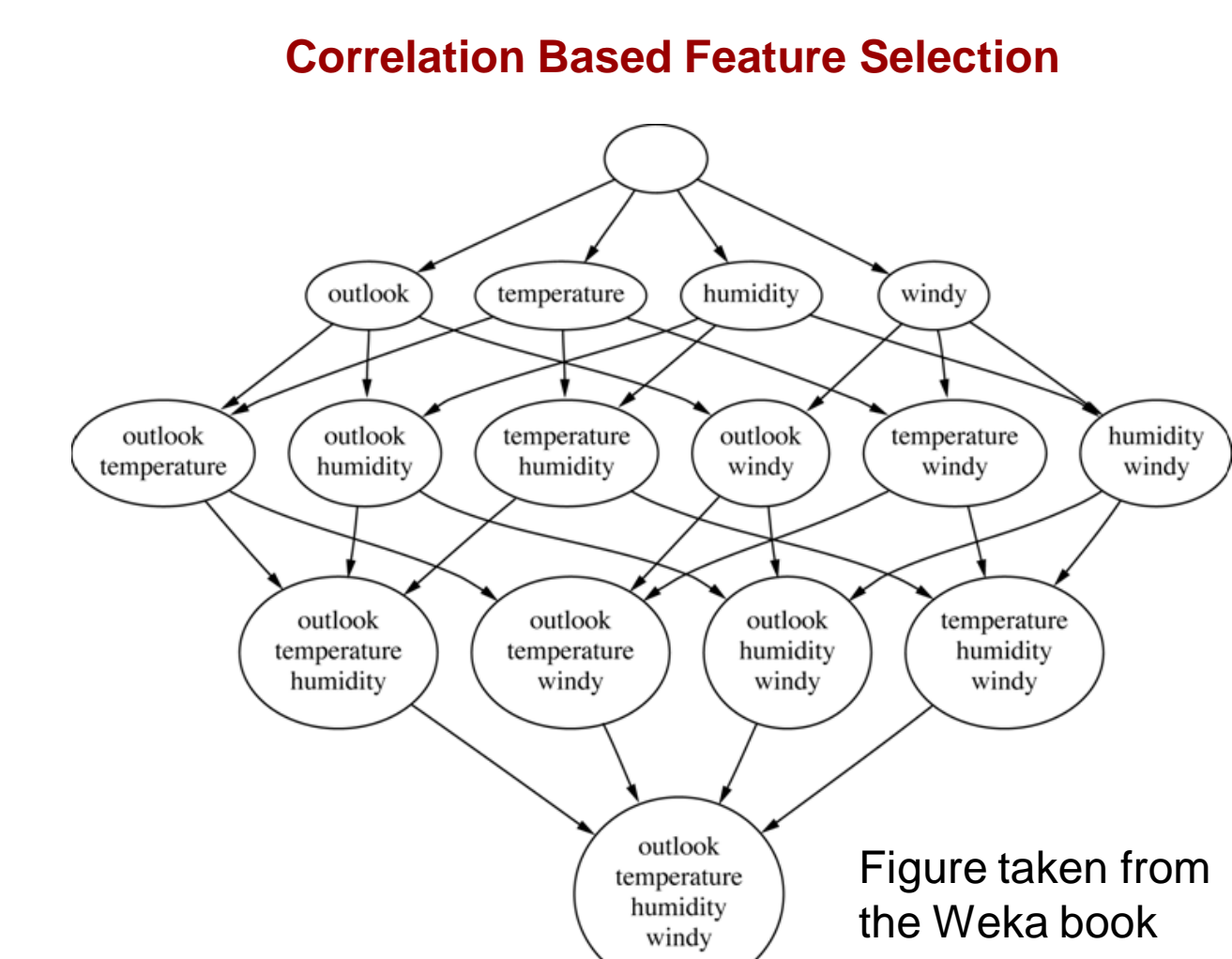
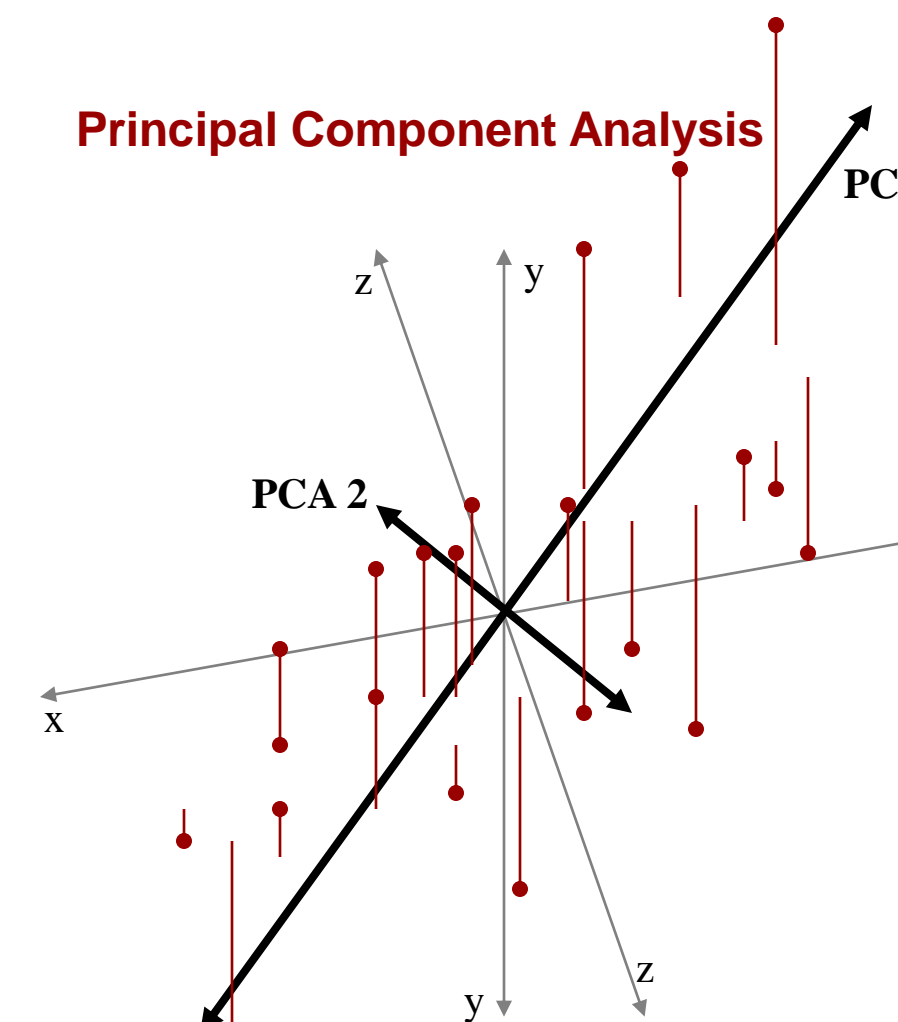
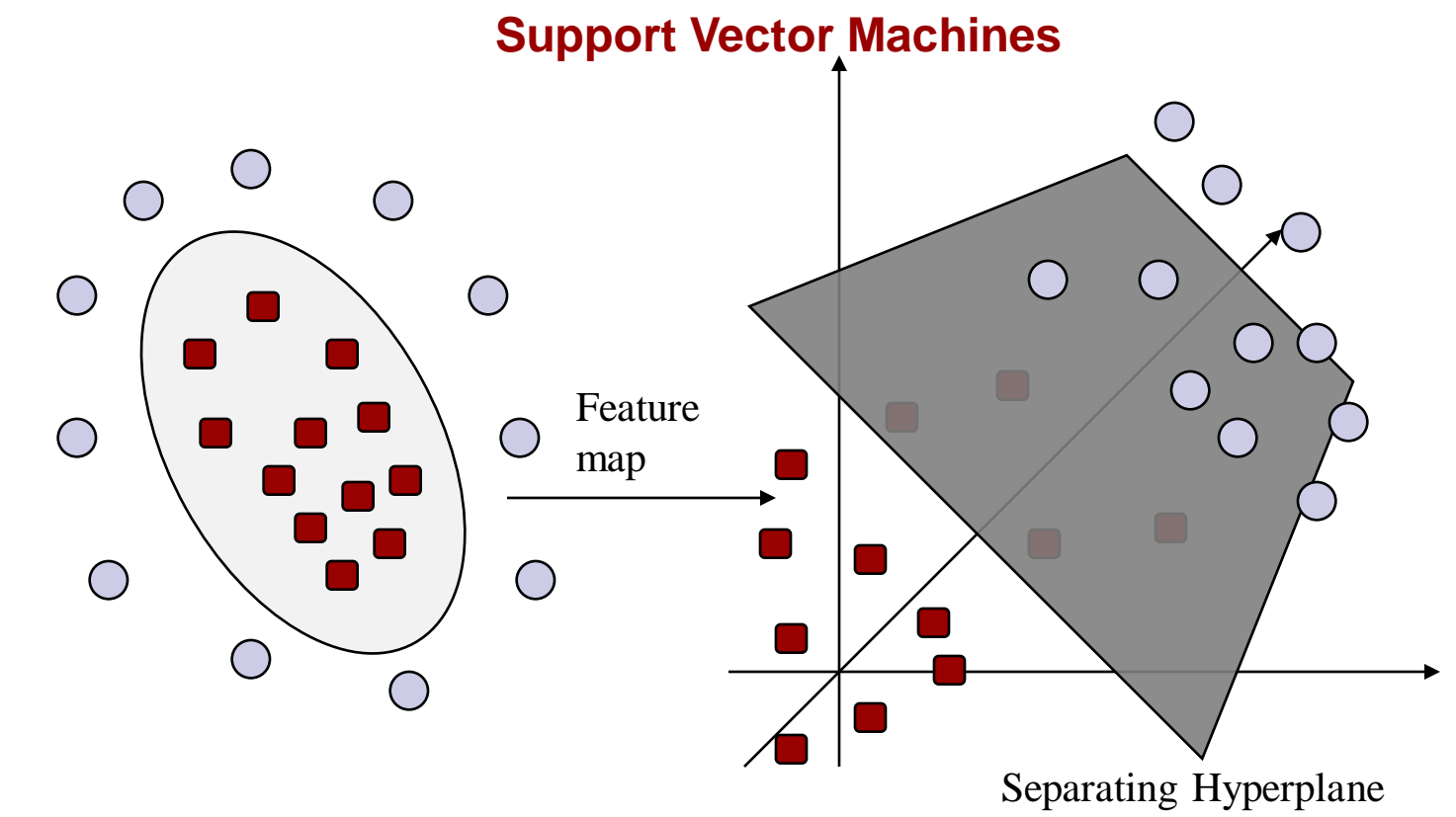
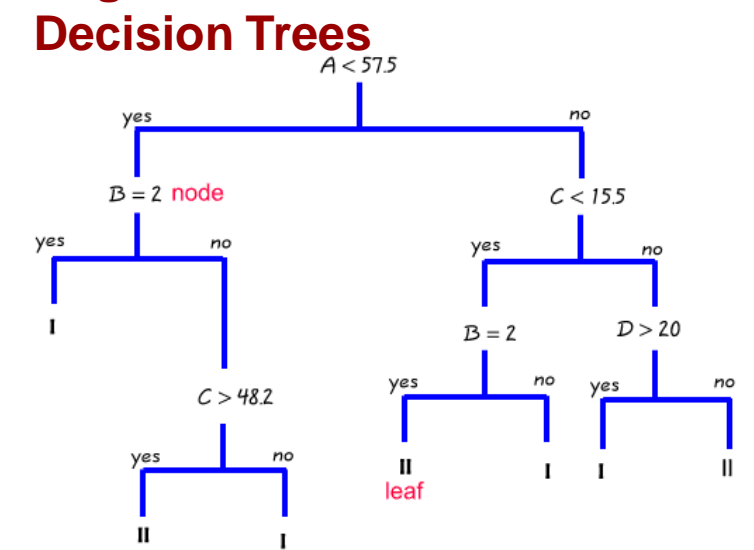
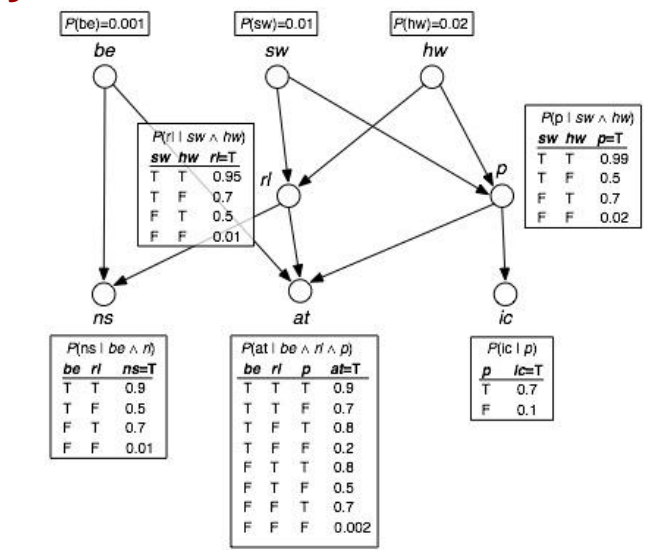


Figure taken from the Weka book

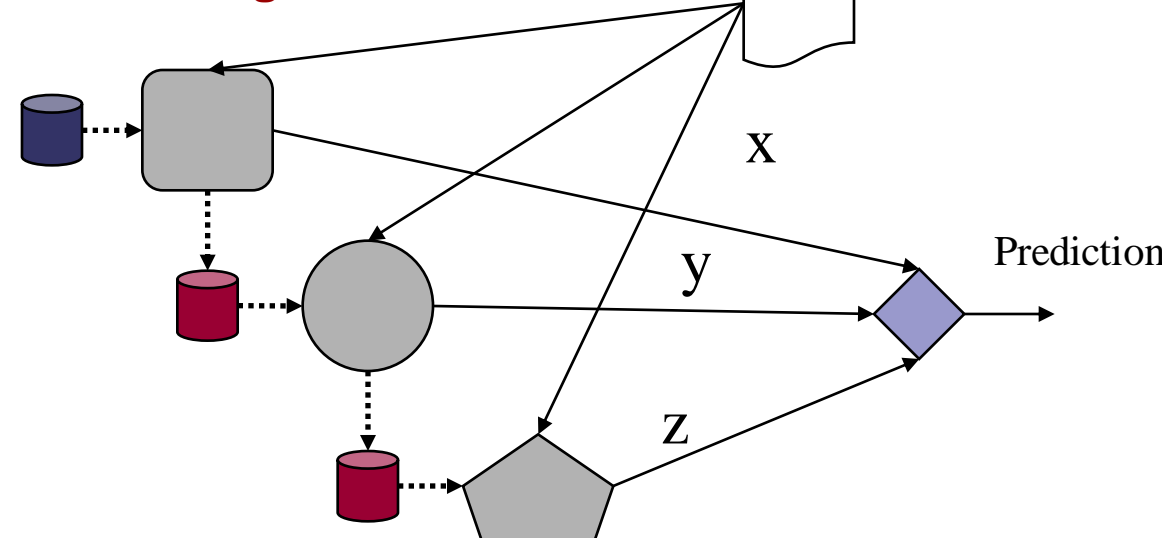
3. Use machine learning algorithms

Naïve Bayes and General Bayesian Models

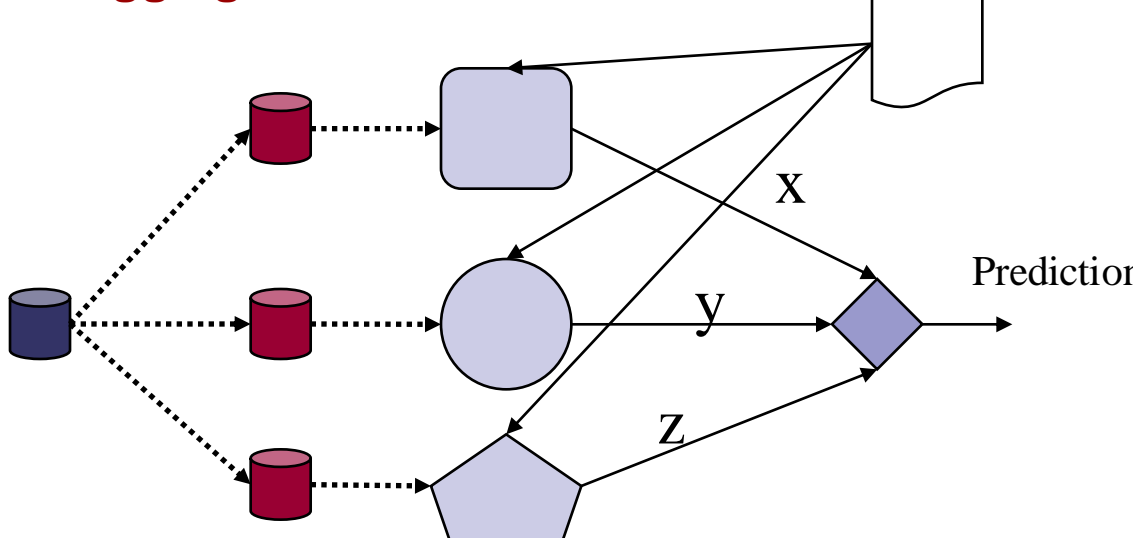
Regression Trees, Model Trees, Decision Trees



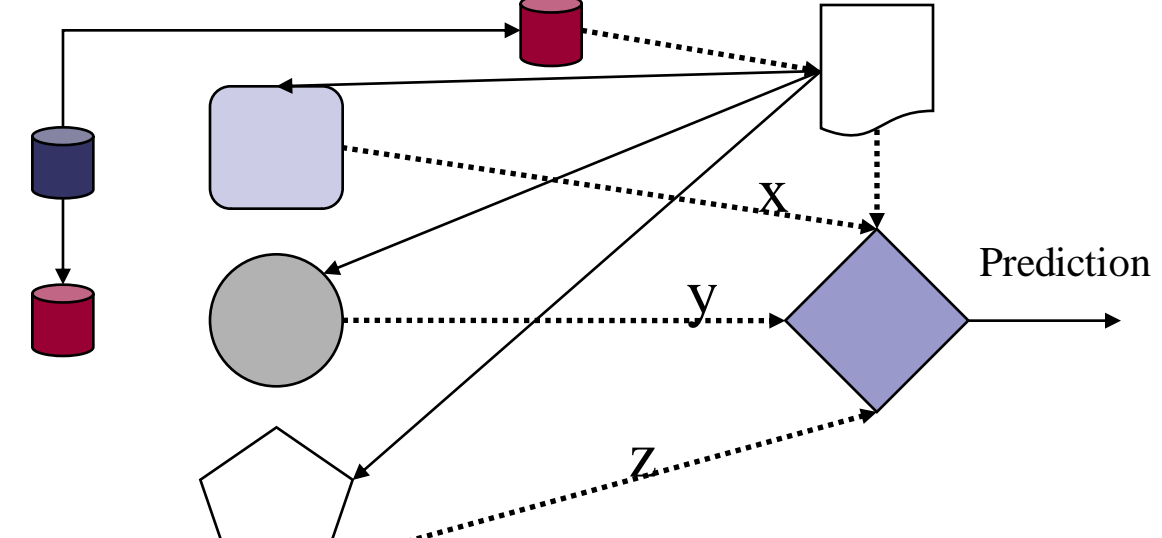
Boosting



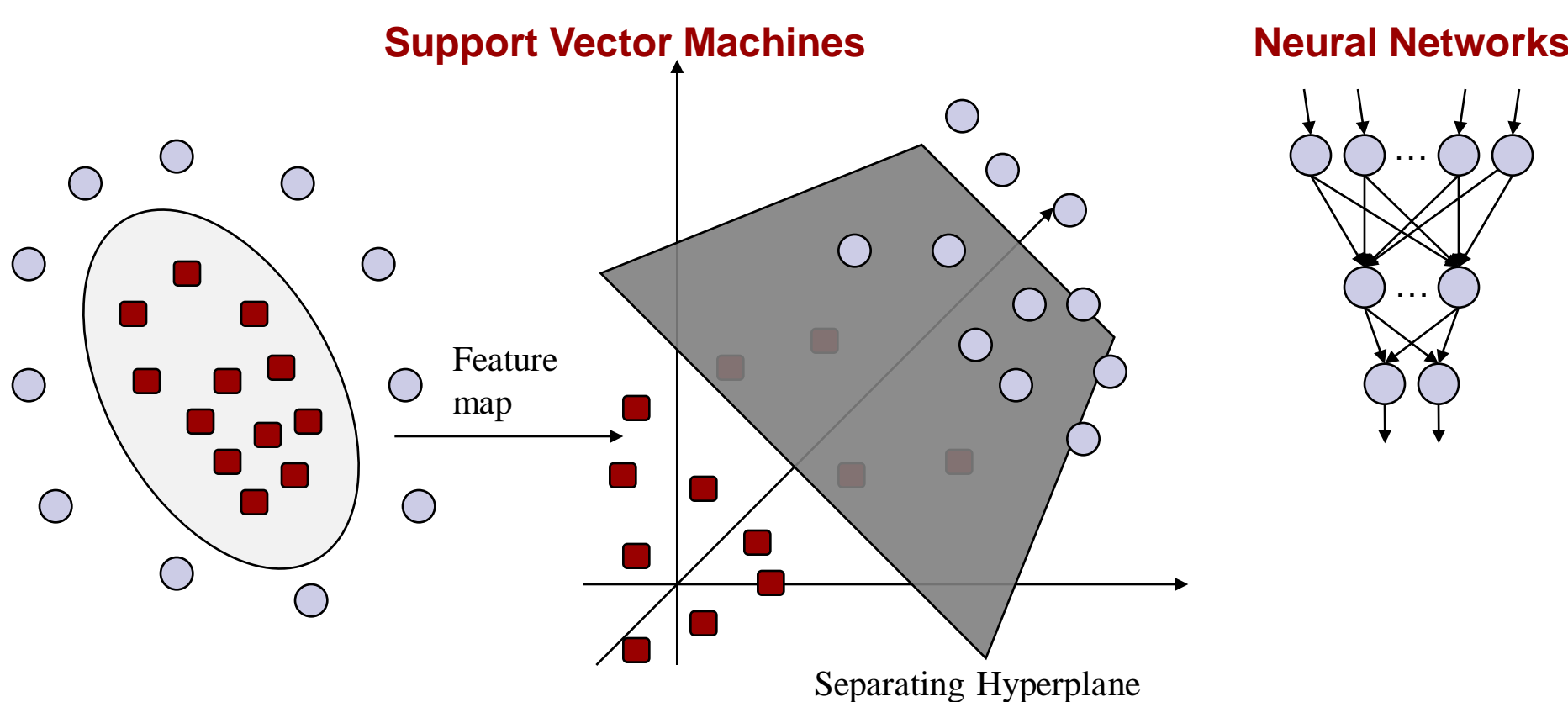
Bagging



Stacking

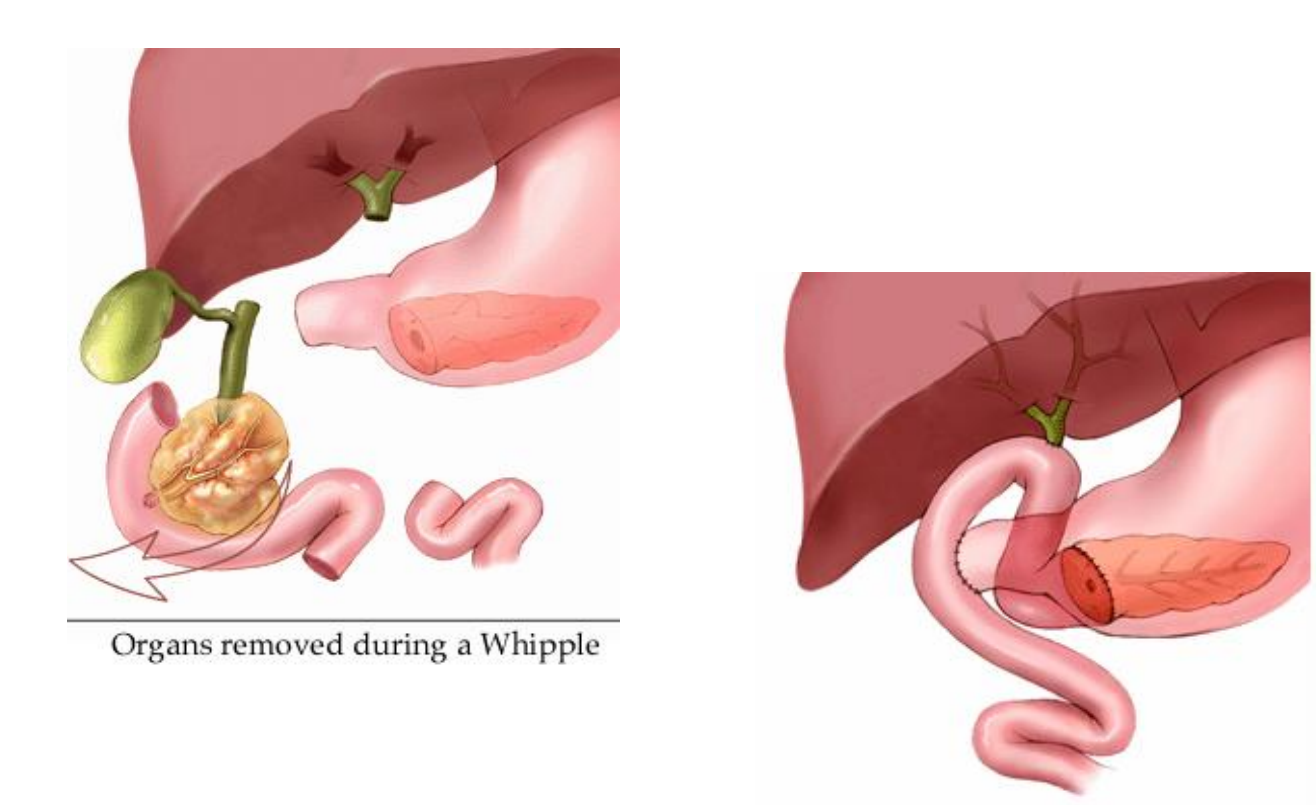


5. Predict patient longevity, quality of life, and provide information factoring into patient care decisions, like surgical removal of tumors



Error-correcting Output Codes

class	class vector	error-correcting class vector
car	1 0 0 0	1 1 1 1 1 1 1 1
chicken	0 1 0 0	0 0 0 0 1 1 1 1
crackers	0 0 1 0	0 0 1 1 0 0 1 1
camera	0 0 0 1	0 1 0 1 0 1 0 1



Most common anatomy after Whipple