www.ispor.org

ISPOR

Improving healthcare decisions

Statistical Learning: What is it? How does it relate to machine learning and AI? How is it applied in healthcare?

Presented by:

Harry J. Smolen, MS
President and CEO
Medical Decision Modeling Inc.

Brought to you by the ISPOR Midwest Regional Chapter

January 30, 2019

- ISPOR Membership

- Member Benefits
- World-class, global scientific conferences
- MEDLINE®-indexed publications
- HEOR resources and tools
- Educational and training programs
- Awards and recognition
- Career development

Statistical Learning: What is it? How does it relate to machine learning and artificial intelligence? and How is it applied in healthcare?

presented by Harry J. Smolen
President & CEO, Medical Decision Modeling Inc.

ISPOR Midwest Regional Chapter
Educational Seminar Presentation Series

MEDICAL DECISION MODELING

Background

- Today's talk will address the following points:
  - What is statistical learning (SL)? (13 slides)
  - Common SL methods, including references to real world healthcare applications (8 slides)
  - Statistical learning example (7 slides)
  - Relationship between SL and machine learning (ML), and relationship between ML and artificial intelligence (AI) (4 slides)

What is statistical learning?

$$Y\ quantitative\ response$$

$$p\ predictors\ X_1, X_2, \ldots, X_p$$

$$assume\ relationship\ between\ Y\ and\ X = (X_1, X_2, \ldots, X_p)$$
$$written\ in\ the\ genereal\ form$$

$$Y = f(X) + \ \varepsilon$$

$$f\ is\ an\ unknown\ function\ of\ X_1, X_2, \ldots, X_p$$

$$\varepsilon\ is\ a\ random\ error\ function$$

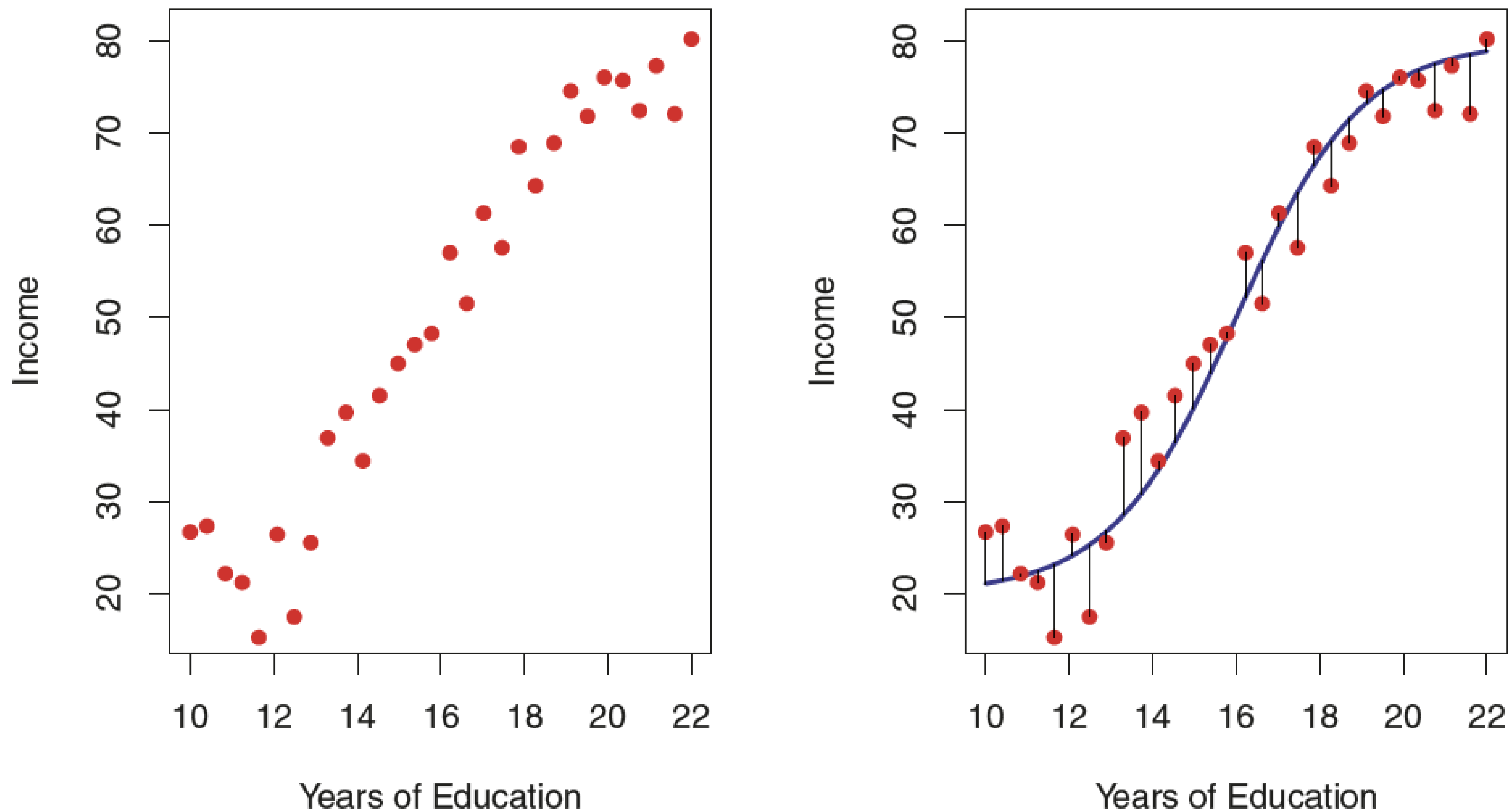
$$\underline{Statistical\ learning\ refers\ to\ methodologies\ for\ estimating\ f^{1}}$$

1. James. An Introduction to Statistical Learning with Applications in R. 2013

What is statistical learning?

$$quantitative\ response\ Y = income^{1}$$

$$p = 1, predictor\ X_1 = Years\ of\ Education$$

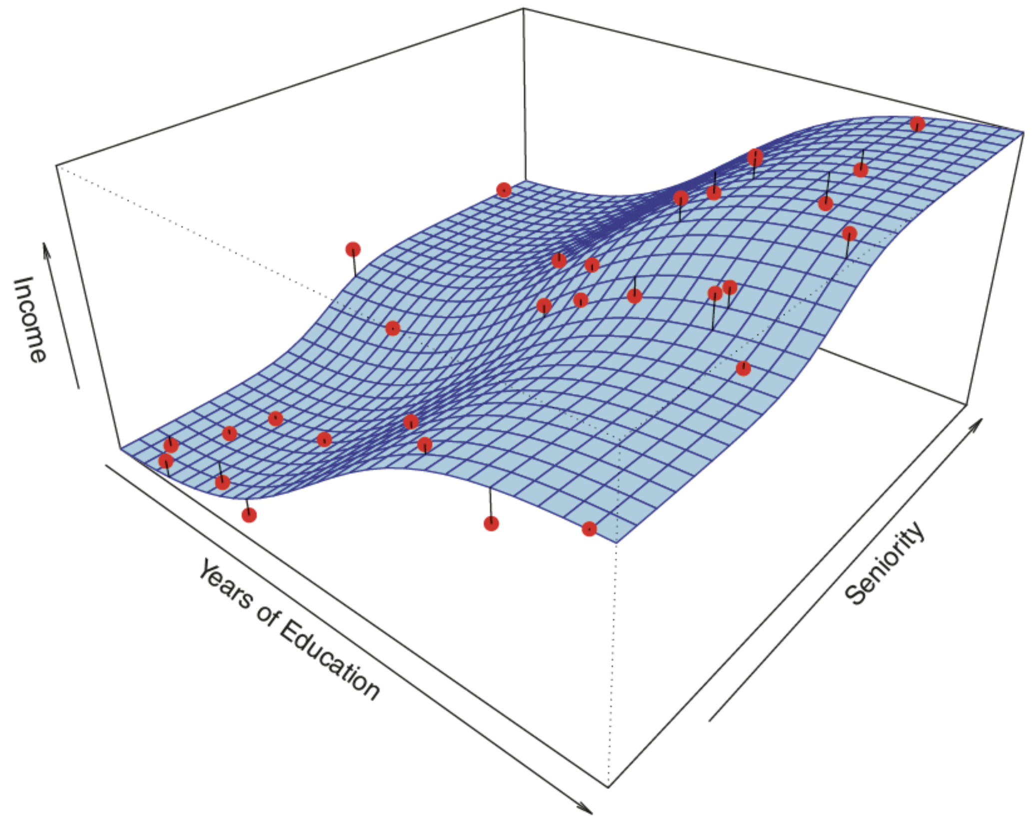


1. James. An Introduction to Statistical Learning with Applications in R. 2013

What is statistical learning?

$$Y = income^1$$

$$p = 2, X_1 = Years\ of\ Education, X_2 = Seniority$$



1. James. An Introduction to Statistical Learning with Applications in R. 2013

What is statistical learning?

- Why estimate *f* ?
  - Prediction
    - Forecasting unobserved outcomes or future behavior
  - Inference
    - Which predictors are associated with the response?
      - Often a small fraction of predictors have a statistically significant association with the response
    - What is relationship between the response and each predictor?
      - Positive or negative relationship?
    - Is relationship between response and each predictor linear? nonlinear?

What is statistical learning?

- To estimate the unknown function *f* we apply statistical learning methods to the training data (see data in above graphs)
- Most statistical learning methods can be classified as:
    - Parametric
    - Non-parametric

What is statistical learning?

- Parametric methods, two-step process
  - First, make assumption about functional form of *f*
    - If assumed *f* is linear $f(X) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$
  - Second, use training data to fit the model
    - In case of linear model we need to estimate parameters $\beta_0, \beta_1, \ldots, \beta_p$
- Assuming a parametric form for *f* greatly simplifies the estimation of *f* because it is easier to estimate a set of parameters (e.g., $\beta_0, \beta_1, \ldots, \beta_p$) than to fit an entirely arbitrary function *f*

What is statistical learning?

- Application of parametric models in health economics
  - NICE recommendations[1] for survival analysis, extrapolating with patient level data
    - Exponential
    - Weibull
    - Gompertz
    - Log-logistic
    - Log normal
    - Generalized gamma
  - Typically, lognormal or gamma distributions for costs
  - Beta or Normal are commonly used to assign patient characteristics

1. Latimer. NICE DSU Technical Support Document 14. 2014

What is statistical learning?

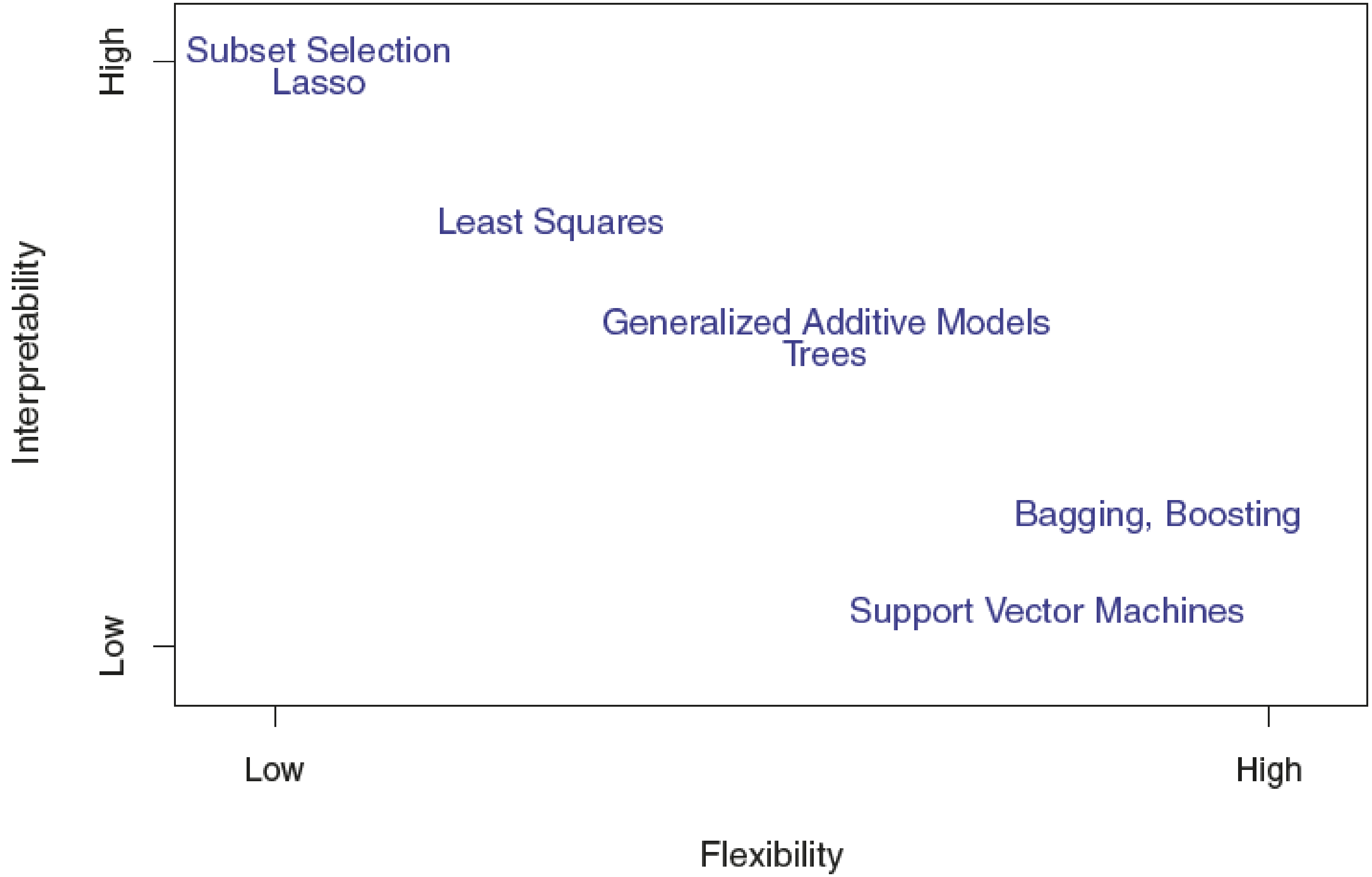
- Non-parametric methods
  - Do not make explicit assumptions about functional form of *f*
  - Seek an estimate of *f* that gets as close to the data points as possible without being too rough or wiggly
  - Major disadvantage: a very large number of observations (far more than is typically needed for a parametric approach) is required in order to obtain an accurate estimate for *f*

What is statistical learning?

- Trade-off between prediction accuracy and model interpretability
  - Restrictive (less flexible) methods can produce a relatively small range of shapes to estimate *f*, e.g., linear methods
  - Other methods, such as the thin plate splines (see 3-D graph above) are considerably more flexible because they can generate a much wider range of possible shapes to estimate *f*
  - Why ever choose a more restrictive method? Interpretability!

What is statistical learning?

Trade-off between flexibility and interpretability using different statistical learning methods[1]



1. James. An Introduction to Statistical Learning with Applications in R. 2013

What is statistical learning?

- Statistical learning can be broadly categorized into:
  - Supervised learning
  - Unsupervised learning

What is statistical learning?

- Supervised statistical learning
  - Involves learning from a training set of data
  - Every point in the training set is an input-output pair, where the input(s) maps to an output
  - Goal is to infer the function that maps between the input and the output, such that the learned function can be used to predict output from future input[1]
  - After selecting a function based on the training set data, this function is validated on a test set of data, i.e., data that did not appear in the training set

1. Poggio. Statistical Learning Theory and Applications. 2012

What is statistical learning?

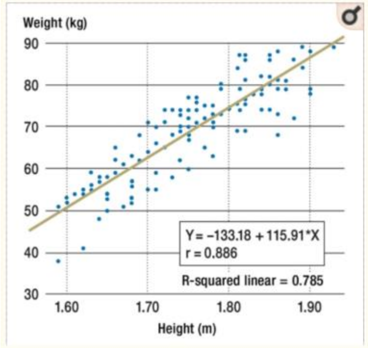
- Unsupervised statistical learning
  - With unsupervised statistical learning, there are inputs but no predefined outcome of interest or label
  - The goal is to learn relationships and structure from the data

Common statistical learning (SL) methods

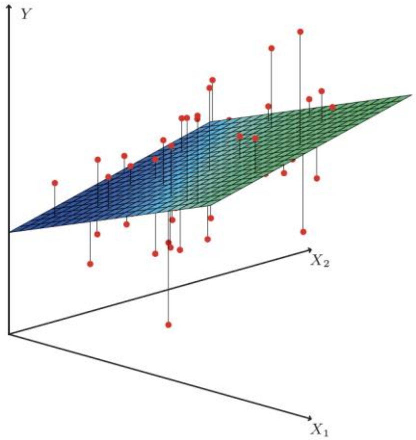
- Two broad types
  - Regression – predicting a quantitative response from one or more predictor variables
    - Expected patient hospital length of stay
  - Classification – predicting a qualitative response from one or more predictor variables by assigning the observation to a category or class
    - Is a patient an acceptable candidate for liver transplantation?

Common SL methods – regression

- Linear regression
  - Simple – one predictor variable[1]



  - Multiple – two or more predictor variables



$X_1 = height$

$X_2 = age$

$Y = weight$

1. Schneider. Dtsch Arztebl Int. 2010

Common SL methods – classification

- Logistic regression
  - Determines probability that response Y belongs to a particular category
    - To predict pancreatic cancer in patients with type 2 diabetes[1]
- Linear discriminant analysis (LDA)
  - Model the distribution of the predictors X separately in each of the response classes, then use Bayes' theorem to estimate $\Pr(Y = k|X = x)$
    - Popular method when more than two response classes
    - Classifying five non-tumorous skin pigmentation disorders[2]

1. Hsieh. Cancer Manag Res. 2018 2. Liang. Comput Biol Med. 2018

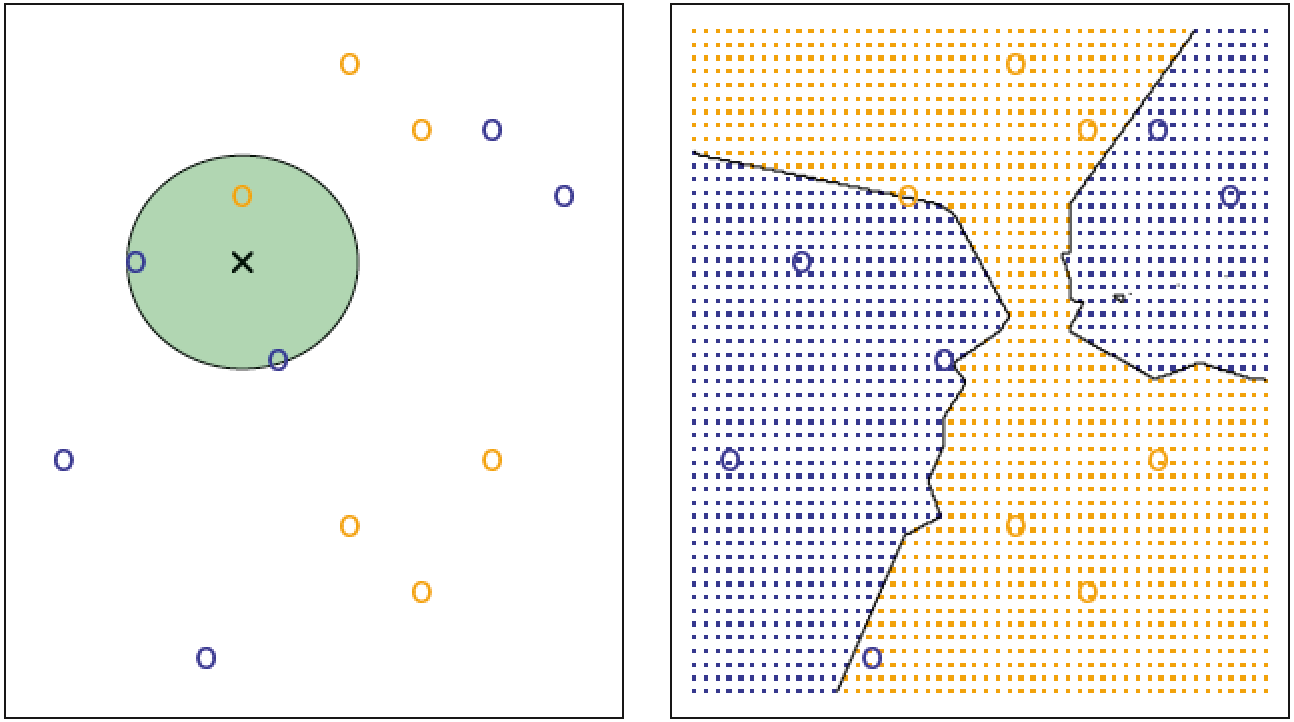
Common SL methods – classification

- K-nearest neighbors
  - With real data we frequently don't know the conditional distribution of Y given X so computing the Bayes classifier (as with LDA) is impossible
  - Estimate the conditional distribution of Y given X, and then classify a given observation to the class with highest estimated probability (see next slide)
  - Differentiate uterine leiomyomas (benign tumors) from uterine sarcomas[1]

1. Nakagawa. Acad Radiol. 2019

Common SL methods – classification

- K-nearest neighbors example[1], $K = 3$



1. James. An Introduction to Statistical Learning with Applications in R. 2013

Common SL methods – classification

- Generalized additive models
  - Framework for extending standard linear models by allowing non-linear functions of each of the variables
    - Estimate the effects of environment factors on cyclist injury severity in automobile-involved bicycle crashes; study classified cyclist injury types as property damage only, possible injury, evident injury, and severe injury or fatality

1. Chen. Accid Anal Prev. 2016

Common SL methods – classification

- Tree-based methods
  - Involve stratifying or segmenting the predictor space into a number of simple regions
  - Rules used to segment the predictor space can be summarized in a tree, hence approaches known as decision tree methods
    - Assist the diagnosis of Alzheimer's disease based on functional magnetic resonance imaging[1]

1. Tripoliti. Conf Proc IEEE Eng Med Biol Soc. 2007

Common SL methods – classification

- Tree-based methods
  - Bagging – use bootstrap to create multiple copies of original training data set, fitting a separate decision tree to each copy, and then combining all of the trees in order to create a single predictive model
  - Boosting – similar to bagging, except that trees are grown sequentially; each tree is grown using information from previously grown trees
  - Random forest – provides an improvement over bagged trees by way of a small tweak that decorrelates the trees

Statistical learning example – genomics[1]

Background

- Unsupervised techniques often applied to analysis of genomic data
- Applied to US NIH's National Cancer Institute NCI60 data set
- Data set has:
    - 64 cancer cell lines (n = 64 observations)
    - 6,830 gene expression measurements per observation (p = 6,830 predictors)

1. James. An Introduction to Statistical Learning with Applications in R. 2013

Statistical learning example – genomics

Problem

- Not predicting a particular output variable
- Based on their gene expression measurements, are there groups or clusters among the cell lines?

Statistical learning example – genomics

Approach

- Hierarchical clustering to estimate groups (clusters)
- Require a dimension reduction method, otherwise based on full data set there are $\binom{6{,}830}{2} = 23{,}321{,}035$ two-dimensional scatterplots, none of which are particularly informative
- Principal component analysis (PCA) as the dimension reduction method
    - PCA is an approach for deriving a low-dimensional set of features from a large set of variables

Statistical learning example – genomics

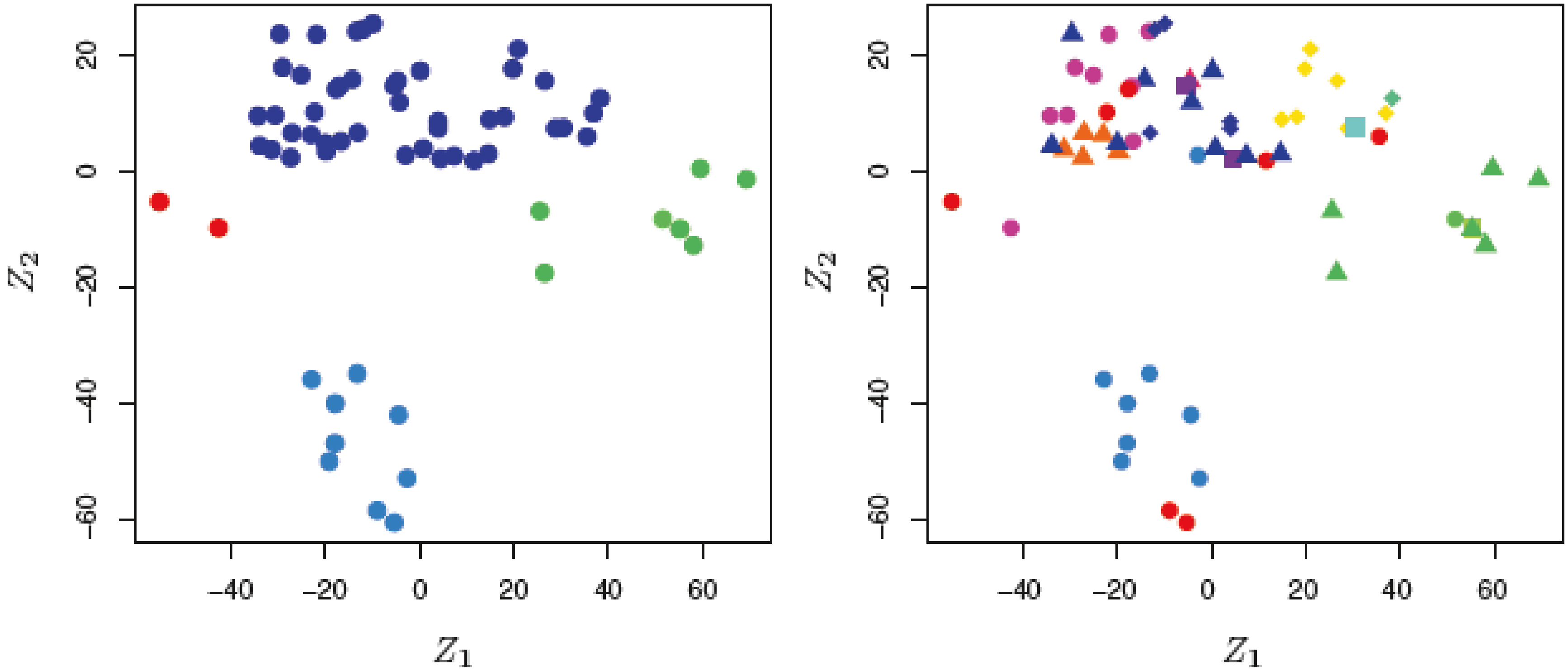
Approach

- Each cell line is associated with one of 14 cancer types (e.g., renal, breast, leukemia, ovarian)
- Consistent with the concepts of unsupervised learning, cell line cancer type is NOT used in performing PCA and clustering
- Cell line cancer type IS used as a check for agreement with unsupervised methods

Statistical learning example – genomics

Results

- Via PCA, 6,830 gene measurements reduced to 2 dimensions Z_1,Z_2 representing first two principal components of data set
- Principal component scores graphed below



Statistical learning example – genomics

Results

- Left panel
  - Each dot represents one of the 64 cell lines
  - Appear to be four groups of cell lines
- Right panel
  - Same as left panel, except each of the 14 types of cancer represented by a unique colored symbol
    - Leukemia, represented by green triangles, all in same cluster
    - Breast cancer, represented by red circles, distributed amongst three clusters
    - Central nervous system (CNS) cancers, represented by orange triangles, grouped closely in top cluster

Statistical learning example – genomics

Conclusions

- Evidence that cell lines with same cancer type tend to cluster in two-dimensional representation
- Though cancer type was not used to produce the left-hand panel, the clustering obtained bears resemblance to some of the actual cancer types observed in the right-hand panel
    - Provides some independent verification of accuracy of the clustering analysis
- Next step would be to examine cell lines within each cluster for similarities in types of cancer, to better understand the relationship between gene expression levels and types of cancer

How does SL relate to ML?

- Statistics draws population inferences from a sample, ML finds generalizable predictive patterns
- In theory, methods from statistics and ML may be used for both data inference and prediction
- However, statistical methods have a long-standing focus on inference, which is achieved through the creation and fitting of a dataset-specific mathematical model
- In statistics, the mathematical model allows the computation of a quantitative measure of confidence that a discovered relationship describes a "true" effect that is unlikely to result from noise[1]

1. Bzdok D. Statistics versus machine learning. Nat Methods. 2018 Apr;15(4):233-234

How does SL relate to ML?

- In contrast, ML concentrates on prediction by using general-purpose learning algorithms to find patterns in often rich and unwieldy data[1,2]
- ML methods are particularly applicable when dealing with “wide data”, i.e., the number of input variables exceeds the number of observations, as opposed to “long data”, i.e., the number of observations exceeds the number of input variables
- ML makes minimal assumptions about the analyzed dataset; ML methods can be effective even when the data are gathered without a carefully controlled experimental design and in the presence of complicated nonlinear interactions
- However, even when ML methods produce convincing prediction results, the absence of an explicit mathematical model can make ML solutions challenging to interpret (i.e., “black box” issue)[3]

1. Bzdok. Classical Statistics and Statistical Learning in Imaging Neuroscience. Front Neurosci. 2017 Oct 6;11:543.
2. Bzdok. Points of Significance: Machine learning: a primer. Nat Methods. 2017 Nov 30;14(12):1119-1120.
3. Bzdok. Statistics versus machine learning. Nat Methods. 2018 Apr;15(4):233-234

How does SL relate to ML?

- Classical statistical modeling was designed for data with a few dozen input variables and sample sizes that would be considered small to moderate in today's "big data" environment
- In classical statistical modeling, the mathematical model fills in the unobserved aspects of the data
- However, as the numbers of input variables and possible associations among them increase, the mathematical model that captures these relationships becomes more complex
- Consequently, statistical inferences become less precise and the boundary between statistical and ML approaches becomes hazier[1]

1. Bzdok. Statistics versus machine learning. Nat Methods. 2018 Apr;15(4):233-234

How does ML relate to AI?

- AI involves machines that can perform tasks that are characteristic of human intelligence, including planning, understanding language, recognizing objects and sounds, learning, and problem solving
- General AI would have all of the characteristics of human intelligence
- Narrow AI exhibits some characteristics of human intelligence, e.g., image recognition[1]
- ML is based on the premise that machines can be built to process data and learn on their own, without constant supervision[2]
- ML is simply a way of achieving AI[1]

1. McClelland. The Difference Between Artificial Intelligence, Machine Learning, and Deep Learning. Dec 4, 2017
2. Mills. Machine Learning vs. Artificial Intelligence: How Are They Different? Forbes. Jul 11, 2018.

Questions?

Harry J. Smolen
President & CEO
Medical Decision Modeling Inc.
3500 DePauw Boulevard, Suite 1037
Indianapolis, IN 46268
USA
(o) 317-704-3800
(m) 317-716-6650
e-mail: smolen@mdm-inc.com