

## techniques to discuss

- neural networks
- support vector machines
- linear regression, including gradient descent, lasso, lars, etc,
- random forests - Breiman
- clustering (see [clustering](#))
- Least squares — Gauss vs Legendre — Stigler discusses
- Chi-squared test — who invented this
- Linear regression — see shift in regression in the 1970s — Hansen lecture1
- PCA
- MCA
- LSA
- Knn
- K-means
- Agglomerative clustering
- Hierarchical clustering
- Decision trees - Breiman
- Regression trees
- Lasso - Efron - Hastie — 2000s
- Lars - Efron 1990s
- CART - Breiman - 1980s
- Linear discriminant function - Fisher -1930s
- Newton's method
- locally weighted regression - worry a little less about having to choose features carefully
- perceptron
- ordinary least squares is just maximum likelihood assuming gaussian errors

## notations & concepts

- the basic  $n$ ,  $x$ ,  $y$ ,  $m$  = number of training examples
- $h$  as the learning output, the hypothesis –
- *features* – feature engineering, feature vector (1 ...  $n$ )
- matrix derivatives – see Ng on gradient descent, Lecture 2/3

## examples

- housing prices
- driving/flying
- prosthetics
- cancer
- spam classification

-

## people

- Pete Norvig, Google, [The unreasonable effectiveness of data] (<http://www.youtube.com/watch?v=...>)  
‘How Billions of Trivial Data Points can Lead to Understanding’  
Sheer volume of data changes success rates  
Shift from rule-based to probability-based: don’t try to work out the rules, but instead  
More data helps
- Hilary Mason
- Alek Kolcz
- Claudia ....
- Mark Hansen
- Chris Bishop
- Kirk L. Wagstaff
- Jimmy Lin
- Andrew Ng
- Hansen? Downloaded all his course materials — really great stuff here —  
all in doc-archive/R/hansen

## clustering

- Everitt, B. (1974). *Cluster Analysis*. London: Heinemann Educ. Books.-
- Hartigan, J. A. (1975). *Clustering Algorithms*. New York:- Wiley.-
- Sneath, P. H. A. and R. R. Sokal (1973). *Numerical Taxonomy*- San Francisco: Freeman.-
- Anderberg, M. R. (1973). *Cluster Analysis for Applications*.- Academic Press: New York.-
- Gordon, A. D. (1999). *Classification*. Second Edition. London:- Chapman and Hall / CRC-
- Murtagh, F. (1985). “Multidimensional Clustering Algorithms”, in
- *COMPSTAT Lectures 4*. Wuerzburg: Physica-Verlag (for algorithmic-details of algorithms used).-
- McQuitty, L.L. (1966). Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. *Educational and Psychological- Measurement*, 26, 825-831.
- Kaufman, L. and Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

This from the ?kmeans in R

- Forgy, E. W. (1965) Cluster analysis of multivariate data: efficiency vs interpretability of classifications- . Biometrics 21, 768–769.
- Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. Applied Statistics 28, 100–108.
- Lloyd, S. P. (1957, 1982) Least squares quantization in PCM. Technical Note, Bell Laboratories. Published - in 1982 in IEEE Transactions on Information Theory 28, 128–137. -MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. . Neyman, 1, pp. 281–297. Berkeley, CA: University of California Press.
- K-means clustering is like the ‘Hello World’ of data. See Hilary Mason’s ‘An Introduction to Machine Learning with Web Data’ <http://shop.oreilly.com/product/0636920017493.do?green=495A8BDC-FF5A-586B-074C-D3C9A9F0A4E5&cmp=af-mybuy-0636920017493.IP>