POLITECNICO DI MILANO
Dipartimento di Elettronica e Informazione

Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
08/09/2011

Answer the following questions identifying the key aspects and try not to exceed the 1.5 page limit per question.

- Use only the 3 sheets provided by the teacher
- ***Write your answers on different sheets according to the question***
- Write your name and Student ID on each sheet you turn in
- English is the official language, however Italian is allowed
- Either pen and pencil are allowed
- No other mean to support yourself is allowed

In case you have special needs (e.g., being graded within a given time) please write it on top of your assignment and ***tell it to the teacher!***.

Question 1: Linear Classification (Answer on sheet 1)

What is a discriminant function and how it is implemented, if it is implemented, in Linear Regression on the Indicator Matrix, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Logistic Regression and in separating hyperplanes methods?

Question 2: Kernel Smoothing (Answer on sheet 1)

What is a Kernel Smoother and how it relates to k-nearest neighbors methods? What kernels are there? How can we use Kernel Smoothers for regression? And how for classification?

Question 3: Clustering (Answer on sheet 2)

Among the different algorithms we have studied, there are two which especially rely on the concept of neighbor. Answer the following questions:

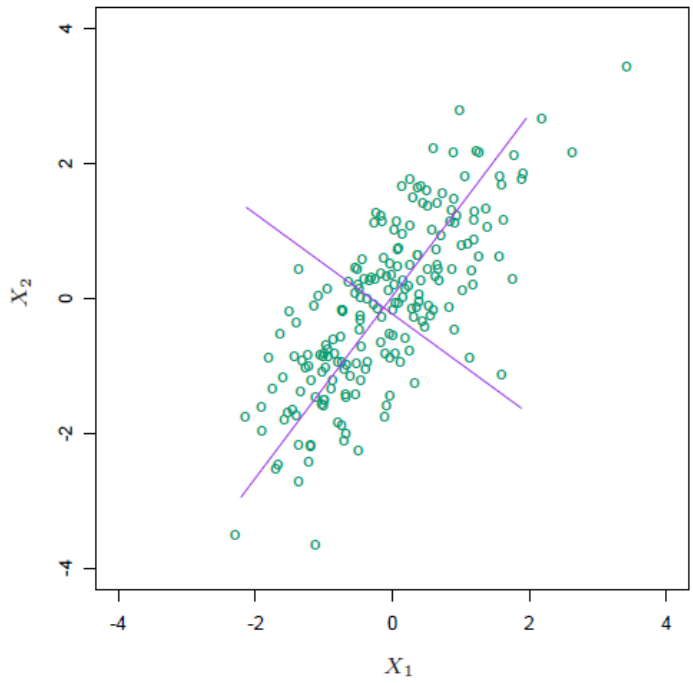
(1) which are these two algorithms?

(2) what are the main differences between the two?

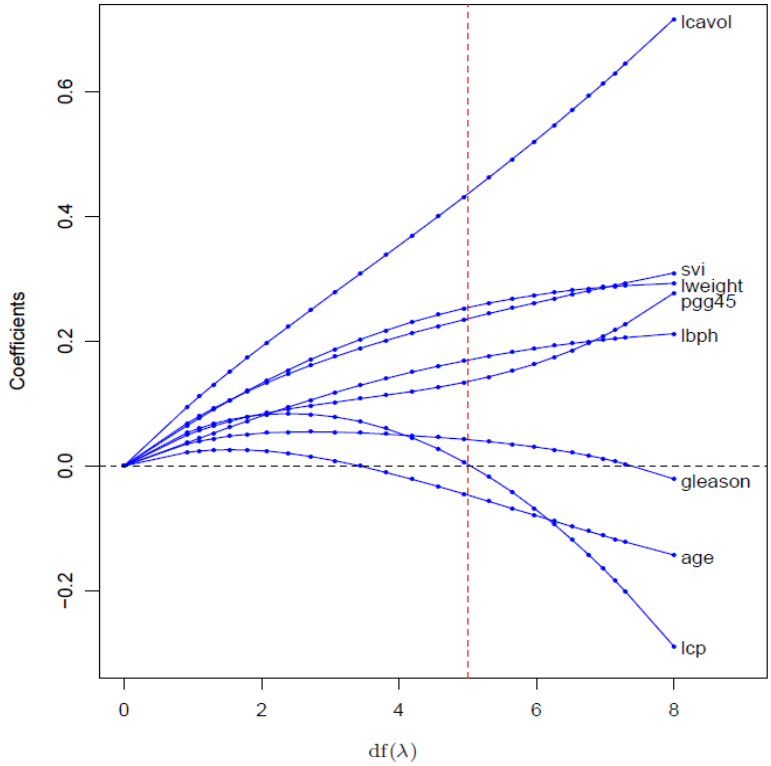
(3) what is the difference between distance and dissimilarity? Does the concept of neighbor need one of them in particular to work?

Question 4: Regression (Answer on sheet 3)

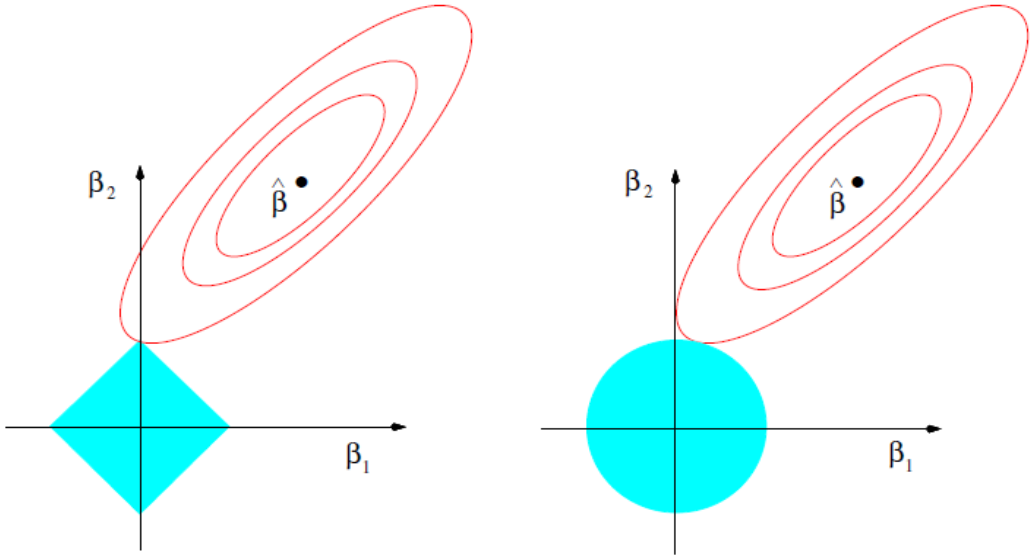
Each of the following images comes from the book "*The Elements of Statistical Learning*". Provide a meaningful caption for each image. Follows the suggestions provided above and motivate your answers. No more than 8 lines for each caption.



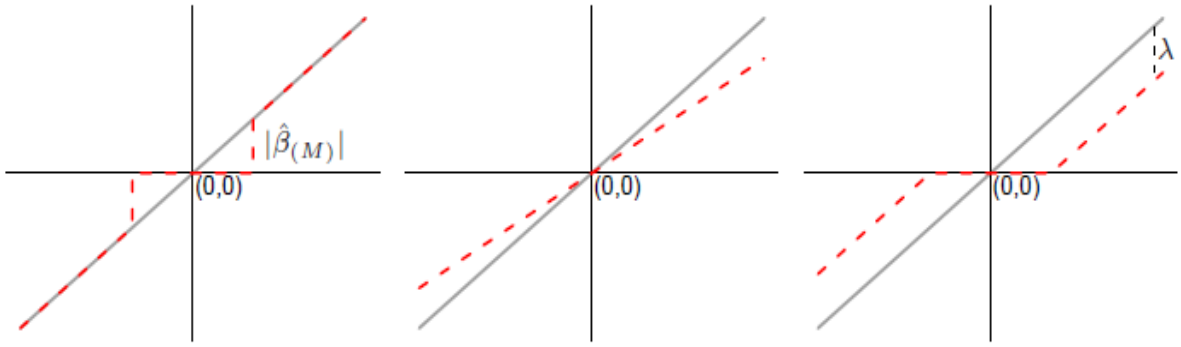
The image above shows some datapoints. The lines are the principal components. How can you obtain them? Between the two, which is the largest principal component? Why? What is its relationship with the distribution (in terms of variance) of the points?



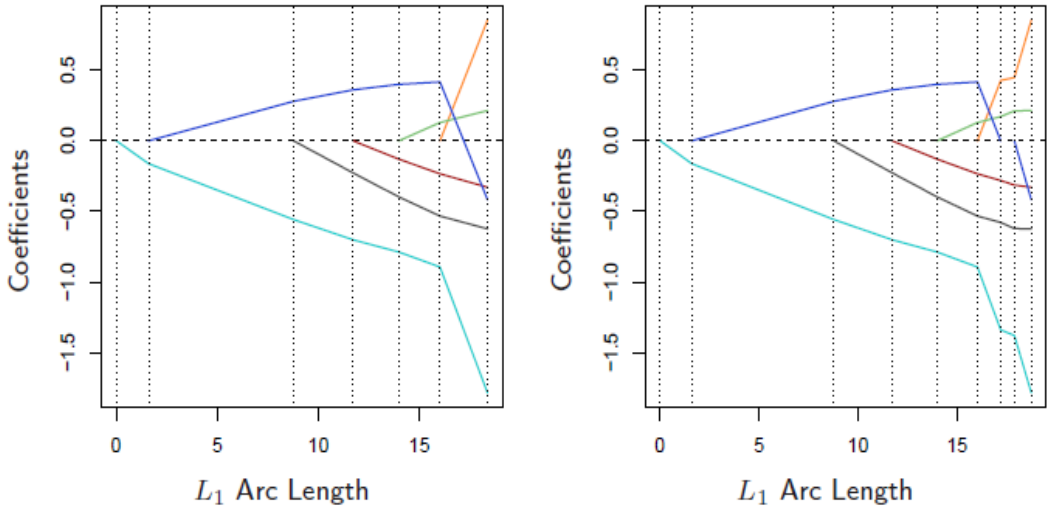
The image above represents the profiles of the regression coefficients of a popular regression method, where lambda is the penalty. Which method is represented? How can you tell? What are the most distictive features of the profile? What is df(lambda)? Which is the relation with lambda? HINT: df(lamba) [chose increases or decreases] as lambda [choose: increases or decreases]



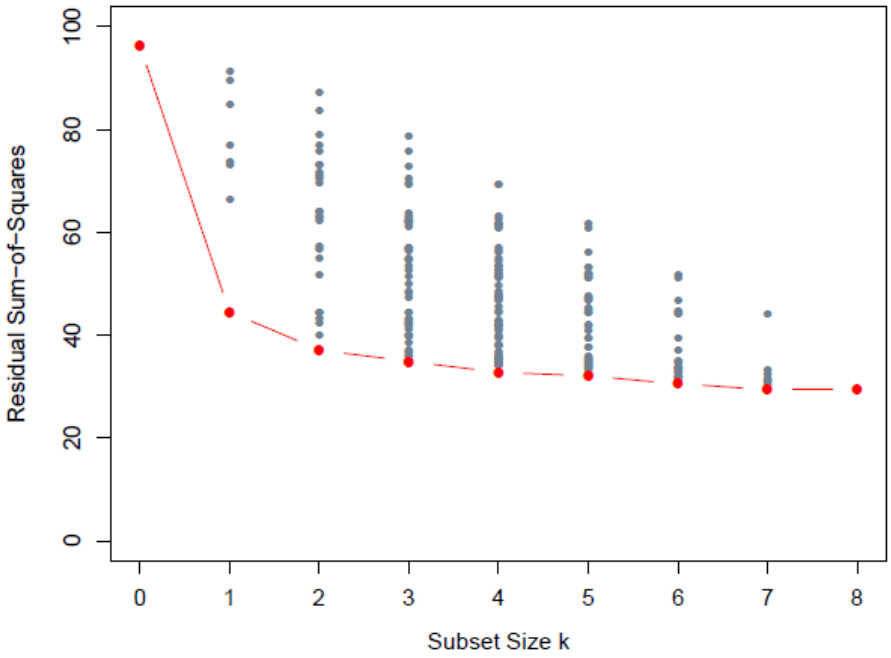
These are the estimation pictures for two popular shrinkage methods. What are they? How can you tell? What are the light blue areas? And the red curves? What is \hat \beta?



The grey lines are the unrestricted estimates for linear regression. The red dotted lines represent three different methods, in case of orthonormal columns X. They are ridge, lasso and best subset. Which is which? Why?



The two panels shows the coefficients of two popular algorithms, LASSO and LAR. Which is which? How can you tell?



Given the cardinality of the subset of variables, what do the grey and red dots represent in the picture above?