EE520, Bollt,

Proj 2, CS, Regression, Inverse Problems, due M Oct 7

This work should be done as a product from each of you, but I do encourage you to help each other extensively – each product finally should be your individual work, codes, words, etc.

The Son of Man (French: *Le fils de l'homme*) is a 1964 painting by the Belgian surrealistpainter René Magritte.

1) 2. Very simple linear regression. Consider the three points (0,1), (1,-1), (2,-2). What is the equation of the line that best fits those three points, in the sense of least squares? State the solution of this problem using methods of SVD. We have seen that the least-squares problem can be stated in terms of the pseudo- inverse which is then in terms of the SVD.

2) In Chapter 3.6 of the provided book excert, there is an example with code, of compressed sensing methods to identify person #7 from a dictionary (theta matrix) built from a library of figures of faces curated as the Yale data set, despite a fancy mustache disguise. Your job here is to work the same data sets, but to recognize person #8 hiding behind a growing apple. For each there should be shown something akin to Fig 3.15-3.16, and a summary plot of quality as a function of apple size. Apple data is provided as a mask.

tiny apple   small apple   Medium Apple   BIG APPLE

a) Adjust the 3.6 code and use the apple data instead of mustache data and make a graph summarizing how apple size corrupts the result and how do you measure that? This is open ended as to how you measure it but there are hints in the book.
b) Do similarly with shot noise corruption of increasing frequency. I.e., akin to fig 3.18.

```
for i=4:14
    S = 2^i;
    support = randsample(n,S);
    x0 = zeros(n,1);   x0(support) = randn(S,1);
    %add this to the figure (column vector) of person 8
    %and then renormalize the pixel values so they are again in the
correct
    %range for a gray scale figure i.e., make your own corrupted data
with
    %this of the person's picture you want to identify by CS.
    %for each of these corrupted figs, just as with the mustache and
the
    %apple
    %try and use CS to identify the person.
end
%make a summary plot of some kind of performance with increasing
corruption
%and give a small discussion and interpretation of these and fig
captions,
%etc.

%Note that this could be interpreted with two different versions of CS.
%One is with equality constraint, argmin ||s||_1 subj y=theta s.
%the other with inequality c
```

2. A. closed form computation.

Let A=[1 2]; y=4. We wish a "good" solution of y=A s. But good is in the eye of the beholder – what you define as good will influence which solution you get.

a. What is the dimension and space of the set s is from. How many s values are there? Draw a picture.

For each of the following optimization problems, since s=(s1, s2), you could make a plot (using plot3d) of J(s) over (s1,s2). You may be able to solve these in closed form with good vector calc skills, or otherwise minimize J for s numerically (as you could plot it). Discuss. When you are done, locate each solution on the picture 2a.

- $J(s)=||As\text{-}y||\_2^2$
- $J(s)=||As\text{-}y||\_2^2+r\,||s||\_2$ (choose r=1). (ridge)
- $J(s)=||As\text{-}y||\_2^2+r\,||s||\_1$ (r=1). (Lasso)

Each of these are underdetermined because of the shape of A. What shape of A would define an overdetermined problem?

3. (Synthetic) Data is supplied regarding allometric scaling relative to basal metabolic rate across many animals of many different sizes and shapes. See paper available for download.

   a. Plot data
   b. Develop the regression model directly to the variables as stated. Develop this regression using the matrix inverse functions format, and do so for a succession of models from linear through polynomial of order q, and q=10. AT what point do you argue there has arisen an overfitting problem? Write out in long hand the role of all matrix computations on paper (mimicking what you saw on the chalkboard), as well as supplying code.
   c. For the same, repeat, by the regularizing concept – ridge regression. Contrast to part b. Again develop the "paper-pencil" description of this process and contrast and interpret results.

Format – Please include statement of question, presentation of solutions with any included mathematical equations, inequalities and descriptive words/narrative AND your interpretation of what you are observing. Include figures, with figure captions and these should also be referred to from within the text. Include any tables, again with captions, and also any code you produced or modified to work these problems. In other words, standard presentation technical writing for technical results.