In this problem set, I am giving more details and suggestions about the least squares and perpendicular projection problems. If you are stuck on parts of the last problem set, I would recommend going forward with this one, and then going back!

Least Squares: Practice setting up

These problems will give a little bit of practice setting up a least squares problem. We won't solve the problems yet: what we will do is write the problem in matrix form, and then convert it to a problem of perpendicular projection. Once we solve the projection problem, we can then apply that solution to these problems. (If you already have a proposed solution to the projection problem, I encourage you to try it here and see how it works out!)

Most of what I ask below is quite close to what I covered in class. Try to do it without notes at first, to see if you can re-create it. Only check the notes if you get stuck!

1. To set the stage, I want to start with fitting a straight line to TWO given data points, so that there is a unique exact solution.
   (a) Suppose we want to fit a straight line $y = mx + b$ to the points $(2, 3)$ and $(6, 5)$. Set this problem up as a matrix equation $A\vec{t} = \vec{y}$. Solve the equation by row-reduction (whichever version of that you learned and/or prefer). Check that your answer is correct. (Of course you could solve this by high school algebra, but I'm practicing the set-up for harder problems.)
   (b) Suppose now we want to fit a straight line $y = mx + b$ to the known points (x_1, y_1) and (x_2, y_2). Do the same process. Your final answer should give formulas for m and b in terms of x_1, y_1, x_2 and y_2. Check that this makes sense with what you know about straight line equations.
   (c) As you know, $y = mx + b$ isn't the only standard form for the equation of a line. Try repeating the first two parts of this question, with the same data points, but trying to fit the equation $Ax + By = C$ to them, with unknown A, B, and C. What is different about the process this time? Does that make sense?
   (d) (Optional) Let's see... how about the form $y - y_0 = m(x - x_0)$? Try fitting the same data points as above, in the first two parts, to this equation, with unknown m, x_0, and y_0. Does your answer make sense graphically?
2. Now, let's repeat the same set-up, but with fitting a straight line to THREE data points. (We won't solve it yet, just set it up.)
   (a) Suppose we want to fit a straight line $y = mx + b$ to the points $(2, 3)$, $(6, 5)$, and $(3, 4)$. Set this problem up as a matrix equation $A\vec{t} = \vec{y}$.
      (i) Try to solve the equation by row-reduction (whichever version of that you learned and/or prefer). How does the row-reduction show you that there is no solution?
      (ii) A reminder about matrix multiplication: check that (however you prefer to do matrix multiplication), that the following is true:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = t_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + t_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix}$$

Interpret this geometrically: what does this say about the set of all vectors $A\vec{t}$, where $\vec{t}$ ranges over all possible inputs? (That is, I'm asking about the set

$$\{\vec{y} \,|\, \vec{y} = A\vec{t} \text{ for some } \vec{t} \in \mathbb{R}^2\}.)$$

(iii) In our least squares problem above, what are the dimensions of the vectors $\vec{t}$ and $\vec{y}$? What dimension spaces does the linear transformation map from and to? Why does that suggest that the equation $A\vec{t} = \vec{y}$ in these dimensions will usually not be solvable for $\vec{t}$, given A and $\vec{y}$? What has to happen for the equation to be solvable after all?

(b) Suppose now we want to fit a straight line $y = mx + b$ to the known points (x_1, y_1), (x_2, y_2), and (x_3, y_3).

(i) Set this up as a matrix equation, as in the previous part.

(ii) Interpret it geometrically as in the previous part. (Nothing really new to do here, I just want you to think it through again.)

(iii) (Optional) Try to solve it by row-reduction. You should end up with a condition on the x_i, y_i which allows it to be solvable. Does the condition makes sense geometrically?

3. Before we go on to first steps in solving the least squares problem, I want to point out that it is not limited to fitting linear relationships. We'll do a lot more of this later, but for now I want to work through one example.

(a) Suppose that we want to fit the data points $(2, 3)$, $(6, 5)$, and $(3, 4)$ to a parabola, $y = ax^2 + bx + c$, where the a, b, and c are unknown. Set this problem up as a matrix equation $A\vec{t} = \vec{y}$. Do you expect that there should exist a solution? If so, why, and if not, why not?

(b) Suppose that we want to fit FOUR data points (x_1, y_1), $\ldots (x_4, y_4)$ to an unknown parabola $y = ax^2 + bx + c$. Set this up as a matrix equation $A\vec{t} = \vec{y}$, and as in the previous questions, discuss what that means: the relevant dimensions, what this looks like geometrically (particularly the set of all $A\vec{t}$), what that says about whether or not you would expect an exact solution.

(c) What will happen if you try to fit a parabola to only TWO data points in this way? Discuss it in matrix terms. (Optionally, it would be interesting to explicitly solve this system, in terms of given data points (x_1, y_1) and (x_2, y_2), and see what the set of all solutions looks like as a family of parabolas.)

Next, I want to establish the claim I made in class: that the set of points $W' = \{\vec{y} \,|\, \vec{y} = A\vec{t} \text{ for some } \vec{t} \in \mathbb{R}^2\}$ forms a linear subspace,[1] that the least-squares solution to $A\vec{t} = \vec{y}$ is the point $\vec{\hat{y}}$ on the subspace W' closest to $\vec{y}$, and that that closest point $\vec{\hat{y}}$ is the perpendicular projection of $\vec{y}$ onto the subspace V'.

4. To start with, I'd like to establish a result that will be needed below, and will also be very useful in future. For this problem, you need to know that in $\mathbb{R}^n$, by

[1] I'm naming it W' in a (most likely futile) effort to get the notation to agree with notation I will use later. Choosing good notation in linear algebra is both important, and a constant headache, as we will see...

definition,[2] the (standard) magnitude of a vector is related to the (standard)[3] dot product by

$$\vec{v} \cdot \vec{v} = \|\vec{v}\|^2 ;$$

that, by definition,

$$\vec{v} \perp \vec{w} \text{ if and only if } \vec{v} \cdot \vec{w} = 0;$$

and that, for any scalars α and α',

$$(\alpha\vec{v} + \alpha'\vec{v}') \cdot \vec{w} = \alpha\vec{v} \cdot \vec{w} + \alpha'\vec{v}' \cdot \vec{w}$$

and

$$\vec{v} \cdot (\alpha\vec{w} + \alpha'\vec{w}') = \alpha\vec{v} \cdot \vec{w} + \alpha'\vec{v} \cdot \vec{w}'.$$

(a) Suppose that $\vec{v}$ and $\vec{w}$ are perpendicular. Prove that

$$\|\vec{v}\|^2 + \|\vec{w}\|^2 = \|\vec{v} + \vec{w}\|^2 .$$

(b) Explain why the previous result is usually called "the Pythagorean theorem".

(c) Explain why this means that

$$\|\vec{v} + \vec{w}\| \geq \|\vec{v}\|$$

and

$$\|\vec{v} + \vec{w}\| \geq \|\vec{w}\|$$

whenever $\vec{v}$ and $\vec{w}$ are perpendicular.

(d) Give an example to show that the inequalities in the previous part may *not* be true if $\vec{v}$ and $\vec{w}$ are not perpendicular.

5. Now, suppose that W' is the set of all $A\vec{t}$, as above. I want to establish that the perpendicular projection $P\vec{y}$ of $\vec{y}$ onto W' is the closest point in W' to $\vec{y}$. This says that the least squares solution $\vec{\hat{y}}$ is the same as the perpendicular projection $P\vec{y}$. More precisely, I want to show that, for any $\vec{z} \in W'$, we have

$$\|\vec{z} - \vec{y}\| \geq \|P\vec{y} - \vec{y}\| .$$

Prove this last statement as follows:

- Draw a diagram, to relate the statement we are trying to prove to a right-angled triangle;
- identify what the sides of this triangle are, as vectors; and
- apply the previous problem to those vectors, choosing $\vec{v}$ and $\vec{w}$ appropriately, to conclude the inequality we're trying to establish here.

Note that your final proof actually won't need the diagram; the purpose of the diagram is to select the right $\vec{v}$ and $\vec{w}$ when you are applying the previous problem.

6. In the above, I took it for granted that W' was a linear subspace. (Can you see where that assumption snuck in?) I already asked you in Problem Set #1 to prove that the set of all vectors representing possible predicted values $\vec{\hat{y}}$ forms a subspace. However, setting it up as matrices as we did above, the statement

[2] In dimension greater than 3, the following formula *defines* what we mean by "length" or "magnitude" of a vector.

[3] Later, we'll see situations where it is necessary to use a different dot product, and we'll see how to do these calculations more generally.

applies to more cases, and is easier to prove. In terms of the above, I am claiming that, given any matrix A, the set of all outputs,

$$W' = \{\vec{y} \,\big|\, \vec{y} = A\vec{t} \text{ for some } \vec{t} \in \mathbb{R}^2\},$$

forms a subspace. (This set is called the *range* of A.) Prove that W' defined above forms a subspace. If you don't remember, recall that you need to prove three things:

(i) $\vec{0} \in W'$;
(ii) if $\vec{z} \in W'$ and $\vec{z}' \in W'$, then also $\vec{z} + \vec{z}' \in W'$; and
(iii) if $\vec{z} \in W'$ and $\alpha \in \mathbb{R}$, then also $\alpha\vec{z} \in W'$.

Projections: Different Approaches

Gee, this problem set is getting long already. At the end of last class, (Class #3, Tuesday February 23), I briefly mentioned three strategies that people were trying, to solve the projection problem. All three of these strategies are worth following up: we're going to use all of them in different ways. I was planning to give hints for the different strategies in problems here, but I will put that off to the next problem set. But please continue working on this problem (by whatever strategy seems promising to you) in the meanwhile!