Example 42. Find the least squares solution to $A\boldsymbol{x}=\boldsymbol{b}$, where

$$A=\begin{bmatrix}4&0\\0&2\\1&1\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}2\\0\\11\end{bmatrix}.$$

Solution. First, $A^TA=\begin{bmatrix}4&0&1\\0&2&1\end{bmatrix}\begin{bmatrix}4&0\\0&2\\1&1\end{bmatrix}=\begin{bmatrix}17&1\\1&5\end{bmatrix}$ and $A^T\boldsymbol{b}=\begin{bmatrix}4&0&1\\0&2&1\end{bmatrix}\begin{bmatrix}2\\0\\11\end{bmatrix}=\begin{bmatrix}19\\11\end{bmatrix}$.

Hence, the normal equations $A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$ take the form $\begin{bmatrix}17&1\\1&5\end{bmatrix}\hat{\boldsymbol{x}}=\begin{bmatrix}19\\11\end{bmatrix}$.

Solving, we immediately find $\hat{\boldsymbol{x}}=\begin{bmatrix}1\\2\end{bmatrix}$.

Check. The error $A\hat{\boldsymbol{x}}-\boldsymbol{b}=\begin{bmatrix}2\\4\\-8\end{bmatrix}$ is indeed orthogonal to $\operatorname{col}(A)$. Because $\begin{bmatrix}2\\4\\-8\end{bmatrix}\cdot\begin{bmatrix}4\\0\\1\end{bmatrix}=0$ and $\begin{bmatrix}2\\4\\-8\end{bmatrix}\cdot\begin{bmatrix}0\\2\\1\end{bmatrix}=0$.

Application: least squares lines

Given data points (x_i,y_i), we wish to find optimal parameters a,b such that $y_i\approx a+bx_i$ for all i.

Example 43. Determine the line that "best fits" the data points $(2,1),(5,2),(7,3),(8,3)$.

Comment. Can you see that there is no line fitting the data perfectly? (Check out the last two points!)

Solution. We need to determine the values a,b for the best-fitting line $y=a+bx$.

If there was a line that fit the data perfectly, then:

$$\begin{aligned}a+2b&=1 &&(2,1)\\ a+5b&=2 &&(5,2)\\ a+7b&=3 &&(7,3)\\ a+8b&=3 &&(8,3)\end{aligned}$$

In matrix form, this is: $\underbrace{\begin{bmatrix}1&x_1\\1&x_2\\1&x_3\\1&x_4\end{bmatrix}}_{\text{design matrix } X}\begin{bmatrix}a\\b\end{bmatrix}=\underbrace{\begin{bmatrix}y_1\\y_2\\y_3\\y_4\end{bmatrix}}_{\substack{\text{observation}\\\text{vector } \boldsymbol{y}}}$ (writing the points as (x_i,y_i))

Using our points, these equations become $\begin{bmatrix}1&2\\1&5\\1&7\\1&8\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}1\\2\\3\\3\end{bmatrix}$. [This system is inconsistent (as expected).]

We compute a least squares solution.

$$X^TX=\begin{bmatrix}1&1&1&1\\2&5&7&8\end{bmatrix}\begin{bmatrix}1&2\\1&5\\1&7\\1&8\end{bmatrix}=\begin{bmatrix}4&22\\22&142\end{bmatrix},\qquad X^T\boldsymbol{y}=\begin{bmatrix}1&1&1&1\\2&5&7&8\end{bmatrix}\begin{bmatrix}1\\2\\3\\3\end{bmatrix}=\begin{bmatrix}9\\57\end{bmatrix}.$$

Solving the normal equations $\begin{bmatrix}4&22\\22&142\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}9\\57\end{bmatrix}$, we find $\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}2/7\\5/14\end{bmatrix}$.

Hence, the least squares line is $y=\frac{2}{7}+\frac{5}{14}x$.

The plot above shows our points together with this line. It does look like a very good fit!

Important comment. In what sense is this the line of "best fit"? By computing a least squares solution the way we do, we are minimizing the error $\boldsymbol{y}-X\begin{bmatrix}a\\b\end{bmatrix}$. The components of that error are $y_i-(a+bx_i)$.

Hence, we see that we are minimizing the **residual sum of squares** $\mathrm{SS}_{\mathrm{res}}=\sum_i[y_i-(a+bx_i)]^2$.

Also see the discussion after the next example (where we swap the role of x and y) as well as the example at the beginning of next class (where we discuss making predictions and why minimizing $\mathrm{SS}_{\mathrm{res}}$ corresponds to minimizing the error of those predictions).

Armin Straub
straub@southalabama.edu

Example 44. (again) Determine the least squares line for the points $(2,1),(5,2),(7,3),(8,3)$.

Solution. Let's repeat the computation we did last class. This time, we let Sage do the actual work for us:

```
Sage] X = matrix([[1,2],[1,5],[1,7],[1,8]]); y = vector([1,2,3,3])
Sage] (X.transpose()*X).solve_right(X.transpose()*y)
```

$$\left(\frac{2}{7}, \frac{5}{14}\right)$$

Here are some intermediate steps to help see what's going on (and that it matches our earlier work):

```
Sage] X.transpose()*X
```

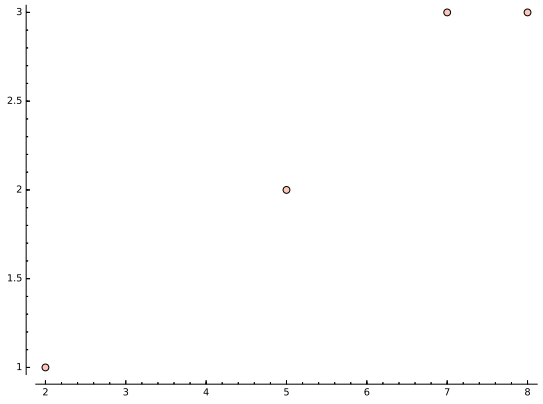
$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

```
Sage] X.transpose()*y
```

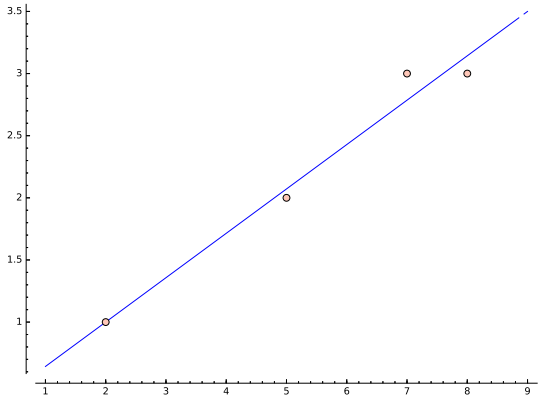
$$(9,\ 57)$$

Let's plot the least squares line $y = \frac{2}{7} + \frac{5}{14}x$ in Sage to marvel at the good fit!

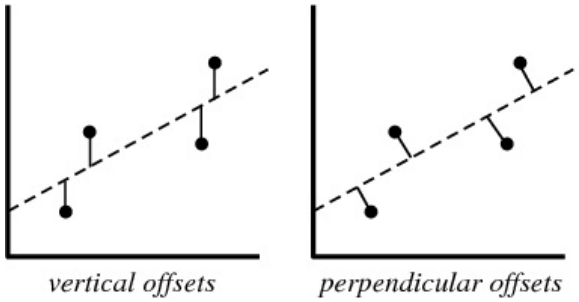
```
Sage] points = [[2,1],[5,2],[7,3],[8,3]]
Sage] scatter_plot(points)
```




```
Sage] scatter_plot(points) + plot(2/7+5/14*x,1,9)
```



Comment. As mentioned earlier, the least squares line minimizes the (sum of squares of the) vertical offsets:



http://mathworld.wolfram.com/LeastSquaresFitting.html

Comment. We get a (slightly) different "best fit" line if we change the role of x and y! Can you explain that?

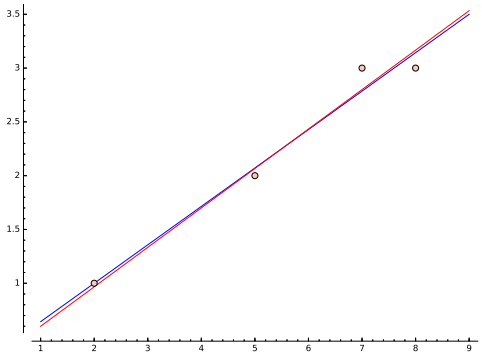
```
Sage] X = matrix([[1,1],[1,2],[1,3],[1,3]]); y = vector([2,5,7,8])
```

```
Sage] (X.transpose()*X).solve_right(X.transpose()*y)
```

$$\left(-\frac{7}{11}, \frac{30}{11}\right)$$

Note that $x = -\frac{7}{11} + \frac{30}{11}y$ is equivalent to $y = \frac{7}{30} + \frac{11}{30}x$.

```
Sage] scatter_plot([[2,1],[5,2],[7,3],[8,3]]) + plot(2/7+5/14*x,1,9) + plot(7/30+11/30*x,
      1,9,color='red')
```



The explanation is that (see pictures at the beginning of this example) we are minimizing vertical offsets in one case and horizontal offsets in the other case.

In linear regression, the relationship between a dependent variable and one or more explanatory variables is modeled. If y is the dependent variable, with x the explanatory variable, then it is natural to minimize the error we make in "predicting y through x" (vertical offsets). See example at the beginning of next class!