Math 312 Jerry L. Kazdan

Least Squares - Weighted

Say we have some data $(t_1,y_1),\ldots,(t_k,y_k)$, where we might think of t as time, and seek a straight line $y=a+bt$ that best fits the data. Ideally, we would like to choose a and b so that

$$\begin{aligned} a+bt_1 &= y_1 \\ a+bt_2 &= y_2 \\ &\vdots \\ a+bt_k &= y_k \end{aligned}.$$

We think of these equations as $A\vec{v}=\vec{y}$, where $\vec{v}=\begin{pmatrix} a \\ b \end{pmatrix}$ and A is the obvious coefficient matrix.

It is highly unlikely that we can find such coefficients a and b. Instead, we seek a and b to minimize the error

$$Q(a,b) := (a+bt_1-y_1)^2+(a+bt_2-y_2)^2+\cdots+(a+bt_k-y_k)^2 = \|A\vec{v}-\vec{y}\|^2. \tag{1}$$

But perhaps we have reason to believe some of the data is more likely to have errors, maybe because it was collected in the middle on the night. This leasd us to weigh the error differently at different times t, say by introducing positive weights $p_1,\ldots, p_k$ and defining the new measure of the error to be

$$E(a,b) := (a+bt_1-y_1)^2p_1+(a+bt_2-y_2)^2p_2+\cdots+(a+bt_k-y_k)^2p_k. \tag{2}$$

If a certain p_j is smaller than the others, then we don't trust that data point as much so it will have less influence on our measurement of the error.

QUESTION: Generalize the Normal Equations to this situations. As a check on your work, observe that is all the weights are equal, $p_1=p_2=\cdots=p_k$, then there is no change in the choice of a and b. In writing your resulty, you may find it useful to introduce the diagonal matrix M whose diagonal elements are $p_1,\ldots,p_k$.

Next, a further generalization of this. The only difference between Q and E is that we have chosen different ways to measure the length of the "error vector" $\vec{z}:=A\vec{v}-\vec{y}$. If we write $\vec{z}=(z_1,\ldots,z_k)$, then in equation (2) we used the following norm to measure the length of the error:

$$|||\vec{z}|||^2 = z_1^2p_1+z_2^2p_2+\cdots+z_k^2p_k.$$

[In a statistical setting, the p_j might represent a probability density and the expression on the left is the variance.] Experience has shown that if one has a norm, it is useful to find a related inner product. Here the following inner product between vectors $\vec{x}$ and $\vec{y}$ is natural:

$$\prec x, y \succ := x_1y_1p_1+x_2y_2p_2+\cdot+x_ky_kp_k. \tag{3}$$

It is eay to verify (by a mental computation) that this has all the properties of an inner product. It is important that the p_j are positive. Our derivation of the *normal equations* for the method of least squares in fact works for any inner product. The only modification needed is that the *adjoint* of a matrix must be defined to fit with the inner produce. For this new inner product, denote the adjoint of a matrix A by $A^\dagger$. The essence is the property that

$$\prec A\vec{x}, \vec{y} \succ = \prec \vec{x}, A^\dagger \vec{y} \succ \qquad \text{for all} \quad \vec{x},\ \vec{y}. \tag{4}$$

QUESTION:

a) With the inner product (3) and $k = 2$, if $A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, compute the adjoint $A^\dagger$.

b) Generalize this to any positive interger k.

Here is a further generalization. A real symmetric matrix M is called *positive definite* if $\langle \vec{x}, M\vec{x} \rangle > 0$ for all nonzero vectors $\vec{x}$ and $\vec{y}$. Define a new inner product by the formula

$$\prec \vec{x}, \vec{y} \succ := \langle \vec{x}, M\vec{x} \rangle.$$

QUESTION:

a) Show that $\prec \vec{x}, \vec{y} \succ$ satisfies all the properties of an inner product.

b) If A is a real matrix, compute its adjoint, $A^\dagger$, with respect to this inner product. The answer will involve both A^* and M.