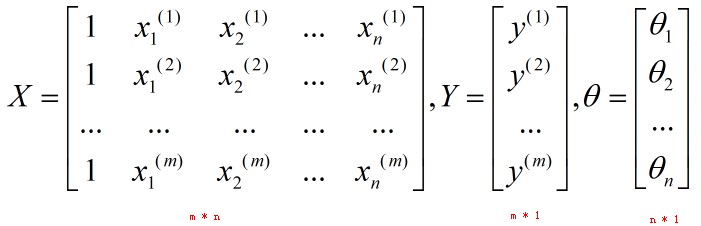
线性回归的计算方法很多，比如最小二乘，梯度下降，今天分享一种矩阵求导计算方法，并且将其与投影联系，可以更加感性的了解线性回归的计算原理。

**Normal Equations 的由来**

假设我们有m个样本。特征向量的维度为n。因此，可知样本为{(x(1),y(1)), (x(2),y(2)),... ..., (x(m),y(m))},其中对于每一个样本中的x(i),都有x(i)={x1(i), xn(i),... ...,xn(i)}。令 H(θ)=θ0+ θ1x1 +θ2x2+... + θnxn，则有



若希望H(θ)=Y，则有

X · θ = Y

我们先来回忆一下两个概念：**单位矩阵** 和 **矩阵的逆**，看看它们有什么性质。

（1）单位矩阵E

AE=EA=A

（2）矩阵的逆A-1

要求：A必须为方阵

性质：AA-1=A-1A=E

再来看看式子 X · θ = Y

若想求出θ，那么我们需要做一些转换：

**step1**：先把θ左边的矩阵变成一个方阵。通过乘以XT可以实现，则有

XTX · θ = XTY

**step2**：把θ左边的部分变成一个单位矩阵，这样就可以让它消失于无形了……

(XTX)-1(XTX) · θ = (XTX)-1XTY

step3：由于(XTX)-1(XTX) = E，因此式子变为

Eθ = (XTX)-1XTY

E可以去掉，因此得到

**θ = (XTX)-1XTY**

这就是我们所说的Normal Equation了。

**Normal Equation VS Gradient Descent**

Normal Equation 跟 Gradient Descent（梯度下降）一样，可以用来求权重向量θ。但它与Gradient Descent相比，既有优势也有劣势。

**优势：**

Normal Equation可以不在意x特征的scale。比如，有特征向量X={x1, x2}, 其中x1的range为1~2000，而x2的range为1~4，可以看到它们的范围相差了500倍。如果使用Gradient Descent方法的话，会导致椭圆变得很窄很长，而出现梯度下降困难，甚至无法下降梯度（因为导数乘上步长后可能会冲出椭圆的外面）。但是，如果用Normal Equation方法的话，就不用担心这个问题了。因为它是纯粹的矩阵算法。

**劣势：**

相比于Gradient Descent，Normal Equation需要大量的矩阵运算，特别是求矩阵的逆。在矩阵很大的情况下，会大大增加计算复杂性以及对计算机内存容量的要求。

**什么情况下会出现Normal Equation，该如何应对？**

（1）当特征向量的维度过多时（如，m <= n 时）

 解决方法：① 使用regularization方式

　　　　　or ②delete一些特征维度

（2）有redundant features（也称为linearly dependent feature）

例如，　x1= size in feet2

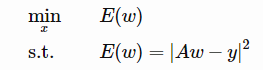
　　　　x2 = size in m2

　　　　feet和m的换算为 1m≈3.28feet所以，x1 ≈ 3.282\* x2, 因此x1和x2是线性相关的（也可以说x1和x2之间有一个是冗余的）

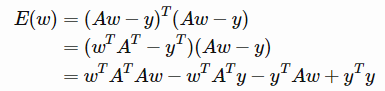
解决方法：找出冗余的特征维度，删除之。

## 计算推导

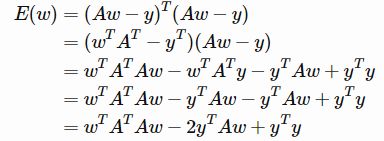
令矩阵A是系数矩阵,且A的列线性独立，一般A的第一列或最后一列是1，用于表示截距，但是这对我们的推导没有任何影响。y是我们的目标结果。现在需要计算一个权重向量w，使得差的平方错误最小，记作:



对E(w)做相关变化



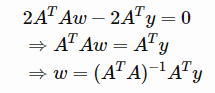
因为为一个1×1的数，所以对其进行转置不变



E(w)是个凸函数，最小值在所有偏导为0的地方，



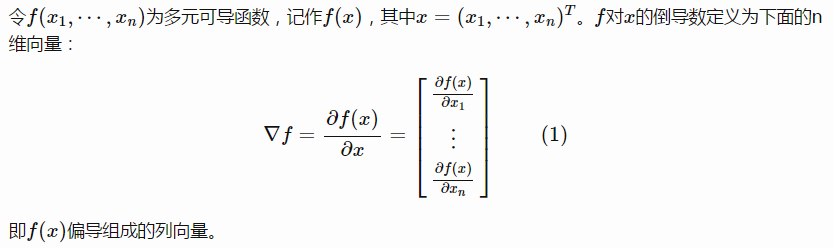
上面的计算使用了向量求导的相关计算，参考[线代随笔11-线性回归相关的向量求导](http://bourneli.github.io/linear-algebra/calculus/2016/04/28/linear-algebra-11-derivate-of-linear-regression.html)。由于A的列线性独立，所以[ATAATA可逆](http://bourneli.github.io/linear-algebra/2016/03/03/linear-algebra-04-ATA-inverse.html)，化简上述公式，



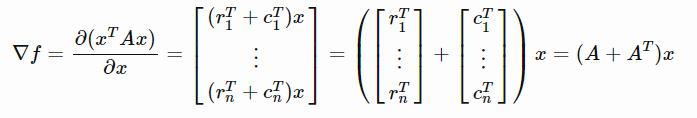
推导完毕！

**相关矩阵导数求解**

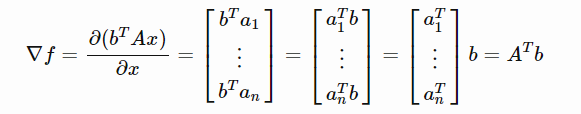
线性回归的计算推导方法有许多，其中有一种使用矩阵运算，涉及到标量对向量的求导，本文主要介绍相关的两个向量求导计算过程











## 线性回归与投影的关系

上述结果与[投影系数](http://bourneli.github.io/linear-algebra/2016/03/05/linear-algebra-05-projection-and-linear-regression.html)的计算公式一致。这不是巧合，线性回归的本质是找到一个线性组合ww，使得因变量yy被由自变量AA的列的线性组合表示。但实际情况，绝大多数是无法找到这种完美的解。那么采取C(A)C(A)中与bb最近的向量作为其近似解。这个最近的向量，通过上面的推导，就是投影系数。可以想象一下三维空间中，直线投影到平面，通过三角关系，可以发现最近的向量是垂直的投影向量。