Python 3 玩儿转机器学习

liuyubobobo

主成分分析

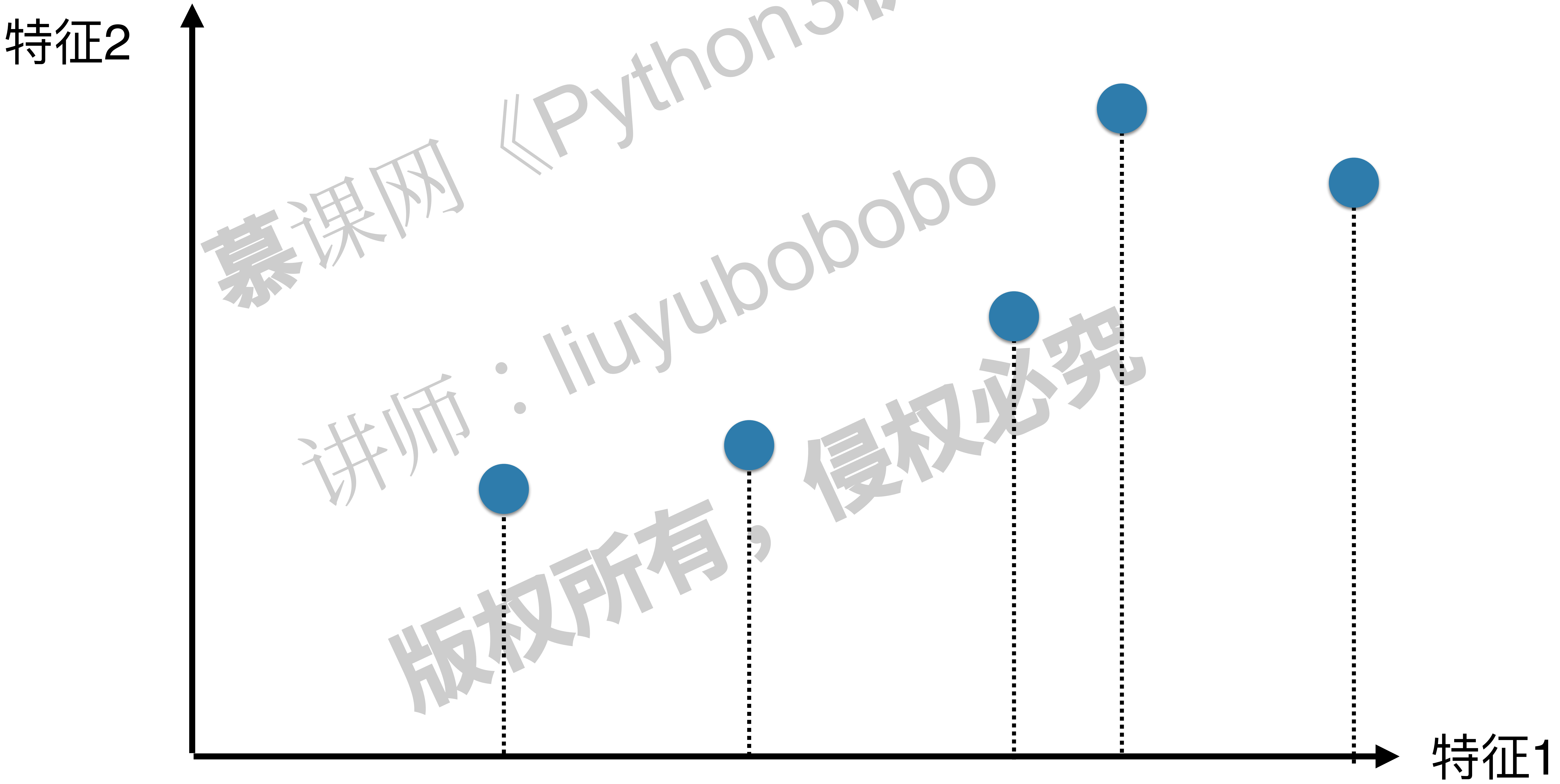
Principal Component Analysis

主成分分析

- 一个非监督的机器学习算法
- 主要用于数据的降维
- 通过降维，可以发现更便于人类理解的特征
- 其他应用：可视化；去噪

主成分分析

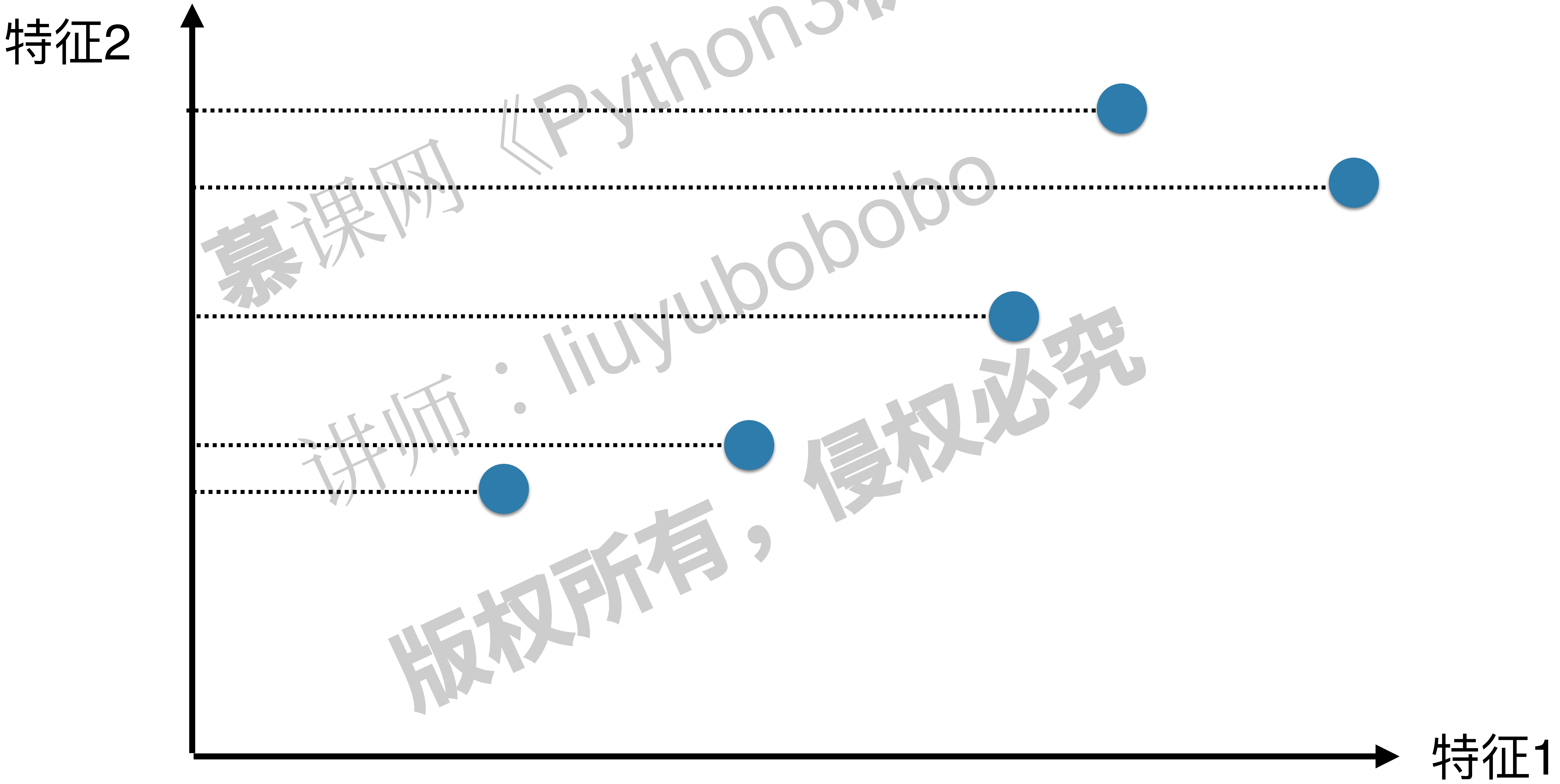
主成分分析



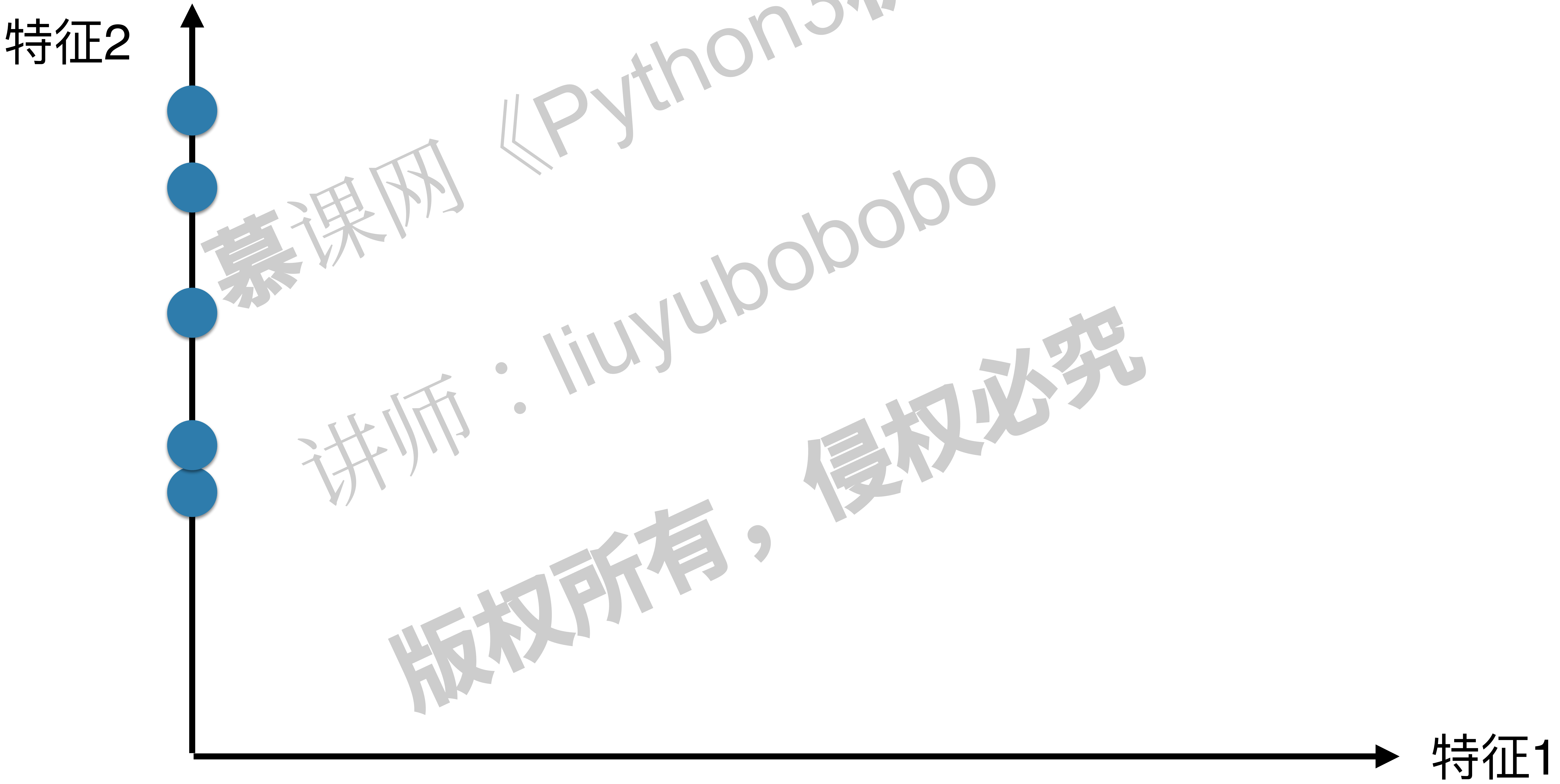
主成分分析

主成分分析

主成分分析



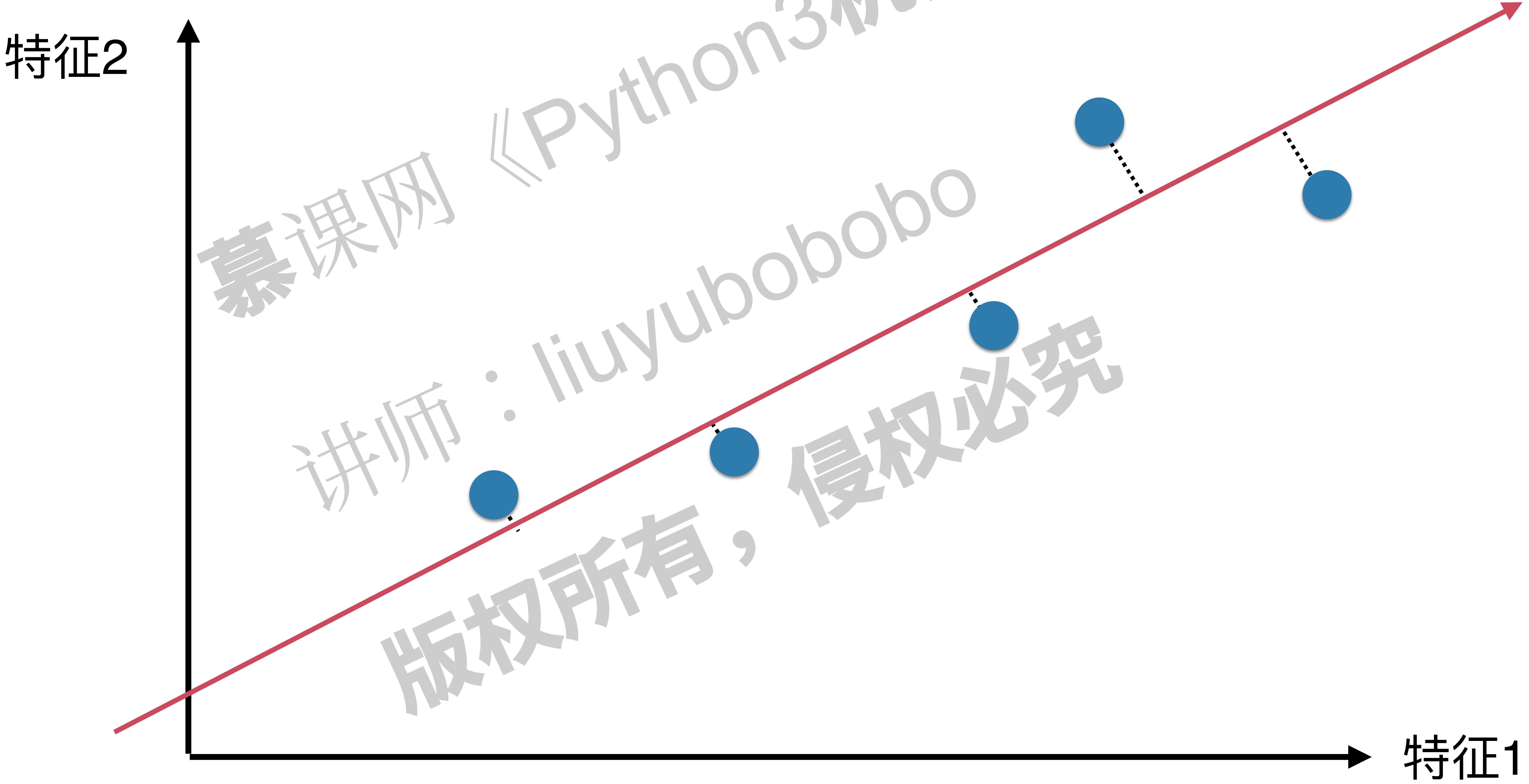
主成分分析



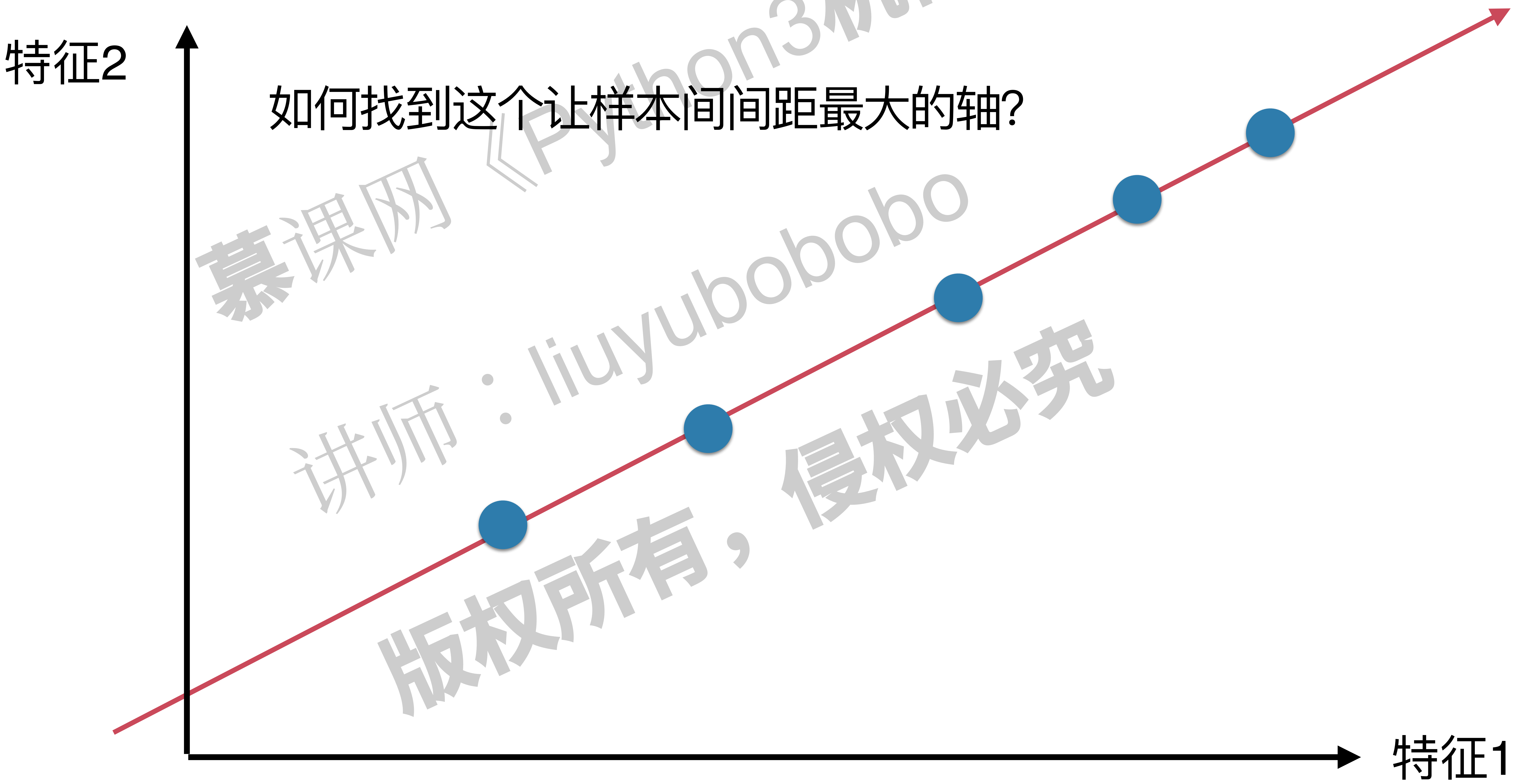
主成分分析

这是最好的方案吗?

主成分分析



主成分分析



主成分分析

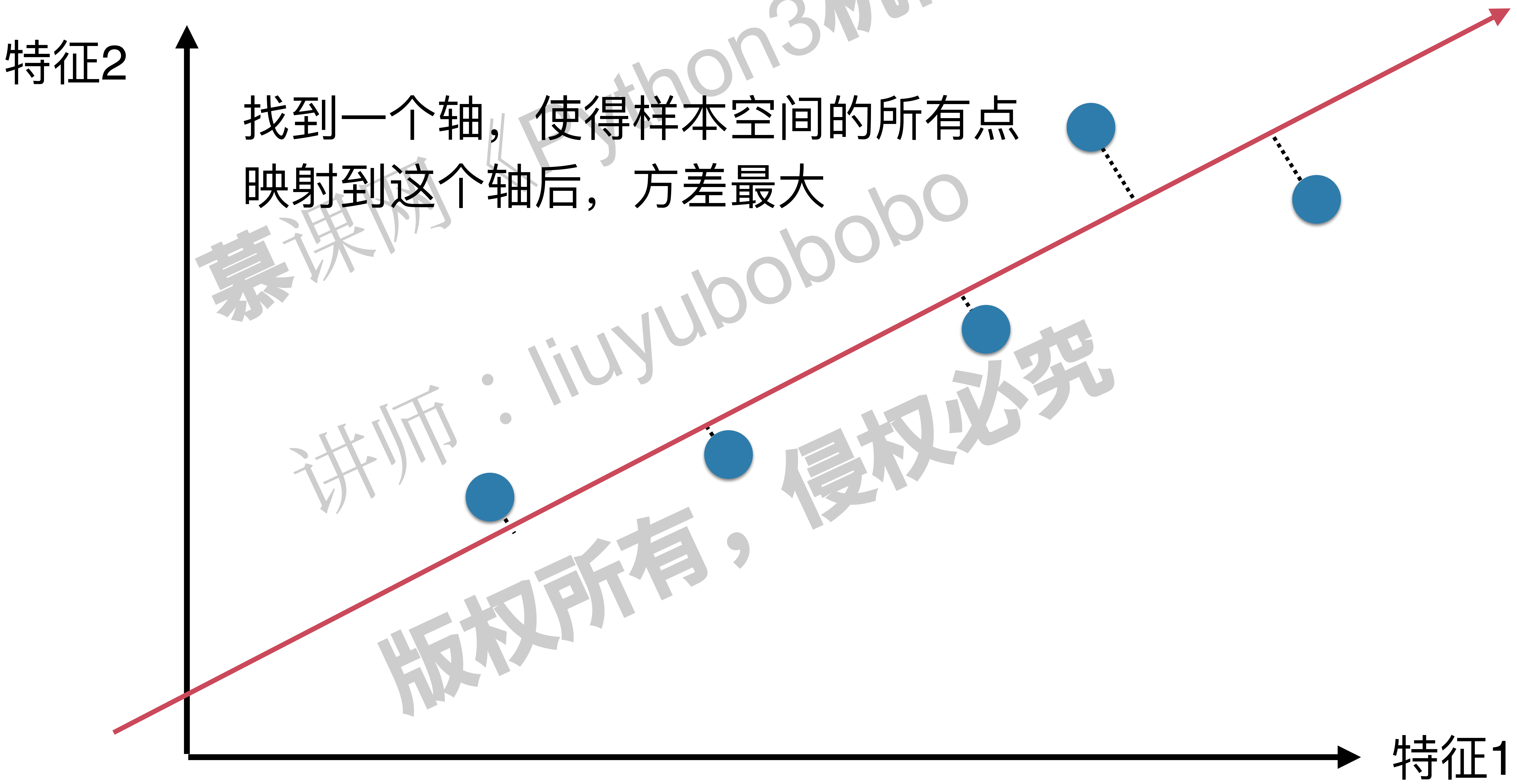
如何找到这个让样本间间距最大的轴?

如何定义样本间间距?

使用方差（Variance）

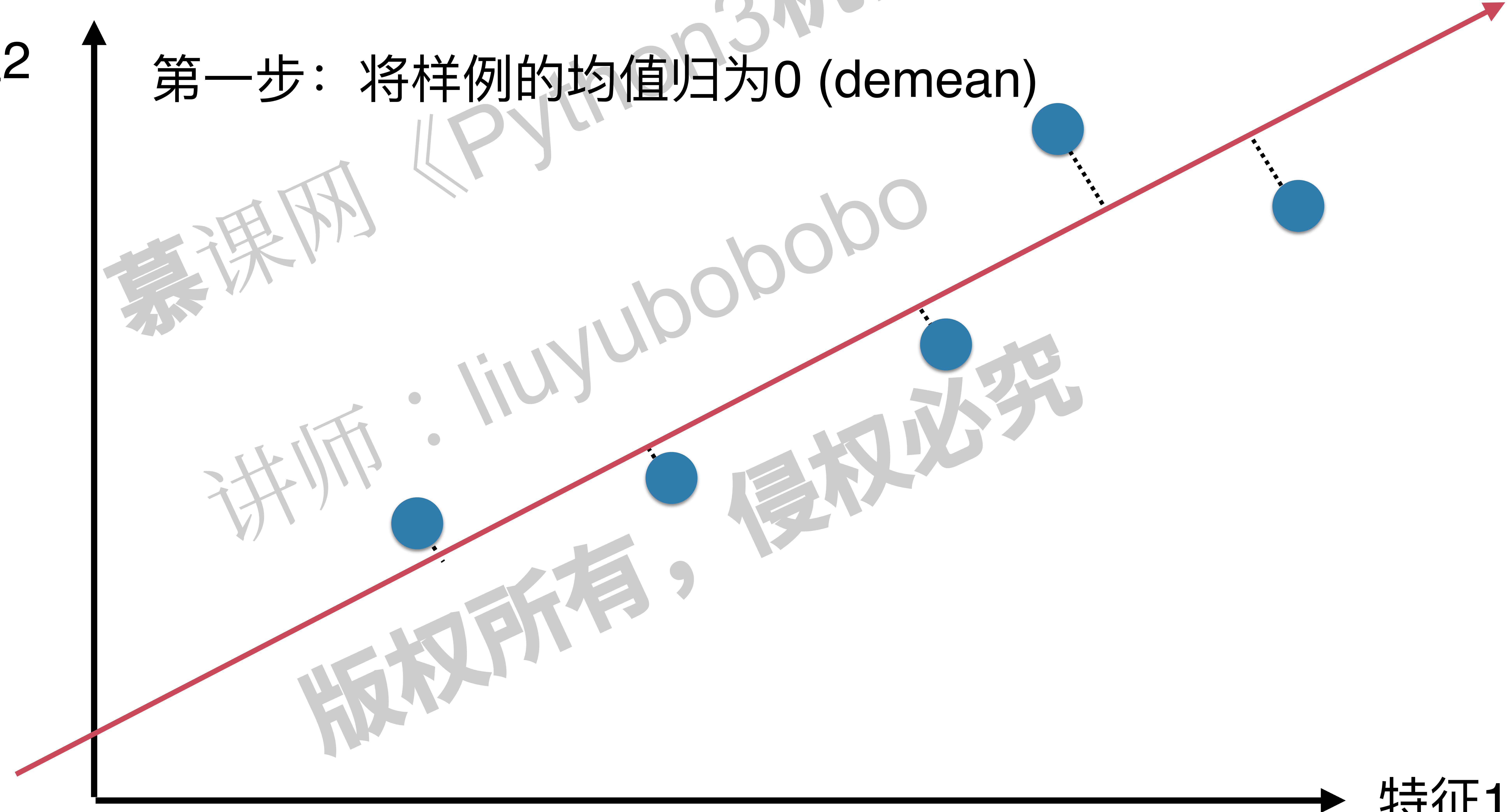
$$Var(x) = \frac{1}{m}\sum_{i=1}^{m}(x_i - \overline{x})^2$$

主成分分析

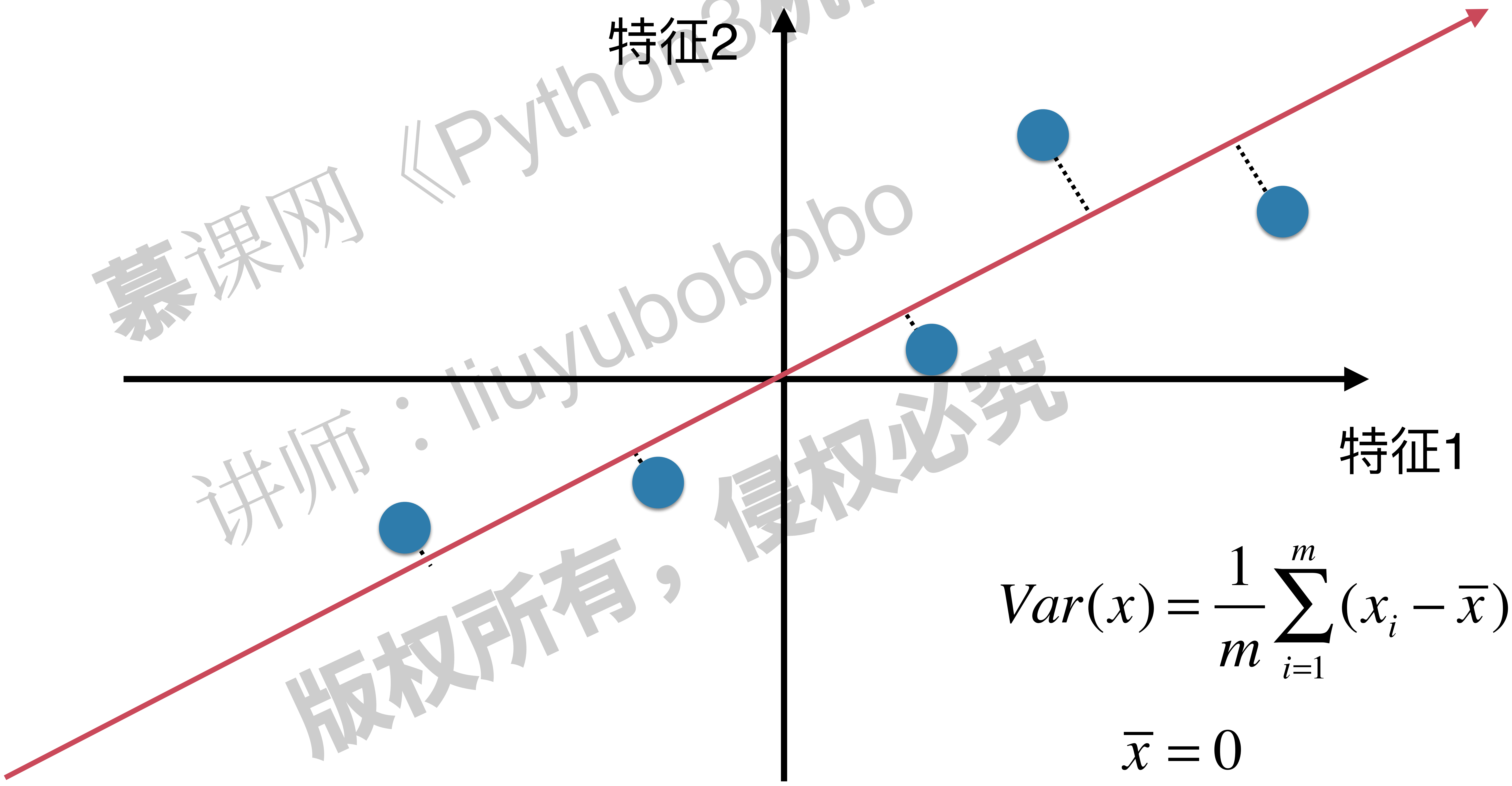


主成分分析

第一步：将样例的均值归为0 (demean)



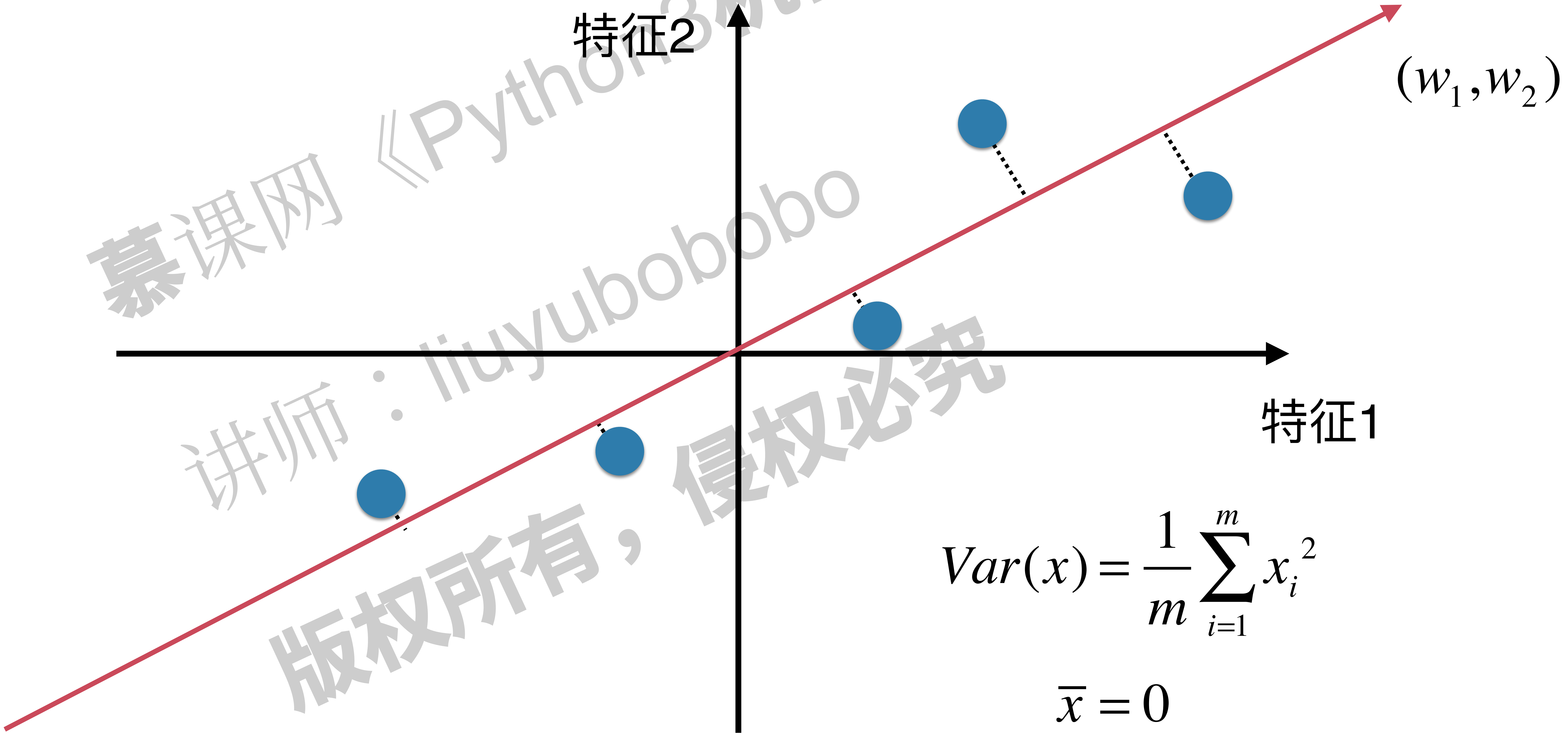
主成分分析



$$Var(x) = \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^2$$

$$\bar{x} = 0$$

主成分分析



$$Var(x) = \frac{1}{m}\sum_{i=1}^{m} x_i^2$$

$$\bar{x} = 0$$

主成分分析

对所有的样本进行demean处理

我们想要求一个轴的方向 w = (w1, w2)

使得我们所有的样本，映射到w以后，有：

$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}(X_{project}^{(i)} - \bar{X}_{project})^2 \quad 最大$$

主成分分析

对所有的样本进行demean处理

我们想要求一个轴的方向 w = (w1, w2)

使得我们所有的样本，映射到w以后，有：

$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}\left\|X_{project}^{(i)} - \bar{X}_{project}\right\|^2 \quad 最大$$

主成分分析

对所有的样本进行demean处理

我们想要求一个轴的方向 w = (w1, w2)

使得我们所有的样本，映射到w以后，有：

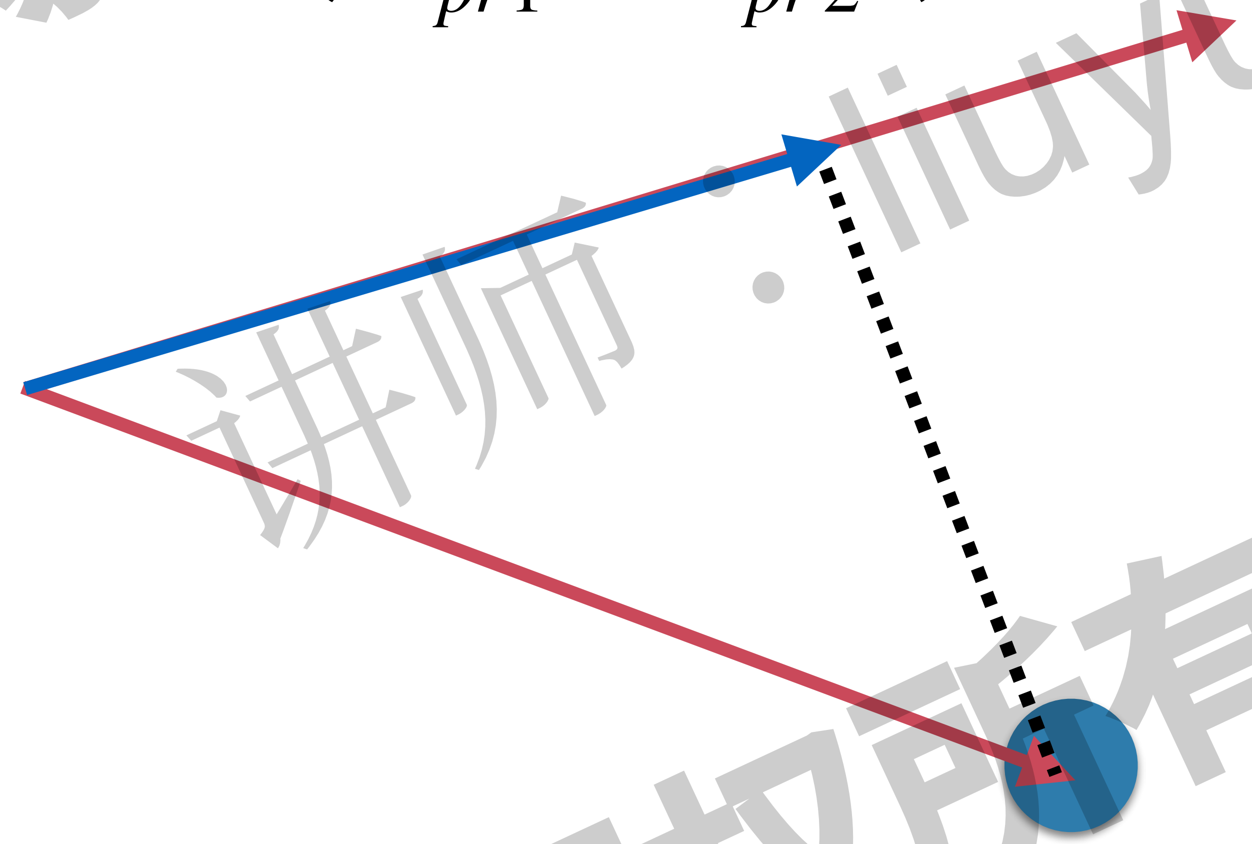
$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}\left\|X_{project}^{(i)}\right\|^2$$ 最大

主成分分析

$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}\left\|X_{project}^{(i)}\right\|^2 \quad 最大$$

$$(X_{pr1}^{(i)}, X_{pr2}^{(i)})$$

$$w = (w_1, w_2)$$



$$X^{(i)} = (X_1^{(i)}, X_2^{(i)})$$

$$X^{(i)} \cdot w = \left\|X^{(i)}\right\| \cdot \left\|w\right\| \cdot \cos\theta$$

$$X^{(i)} \cdot w = \left\|X^{(i)}\right\| \cdot \cos\theta$$

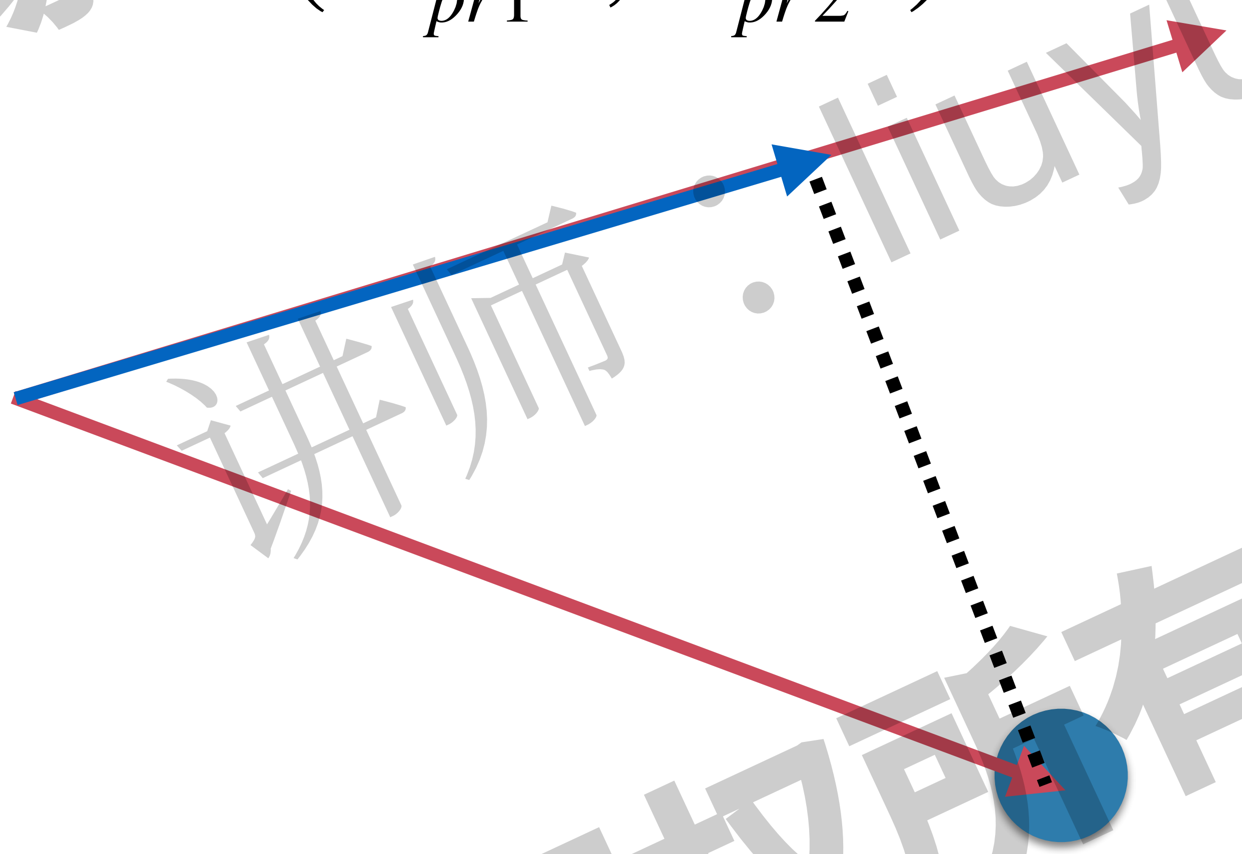
$$X^{(i)} \cdot w = \left\|X_{project}^{(i)}\right\|$$

主成分分析

$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}\left\|X^{(i)} \cdot w\right\|^2$$ 最大

$(X_{pr1}^{(i)}, X_{pr2}^{(i)})$

$w = (w_1, w_2)$



$X^{(i)} = (X_1^{(i)}, X_2^{(i)})$

$$X^{(i)} \cdot w = \left\|X^{(i)}\right\| \cdot \left\|w\right\| \cdot \cos\theta$$

$$X^{(i)} \cdot w = \left\|X^{(i)}\right\| \cdot \cos\theta$$

$$X^{(i)} \cdot w = \left\|X_{project}^{(i)}\right\|$$

主成分分析

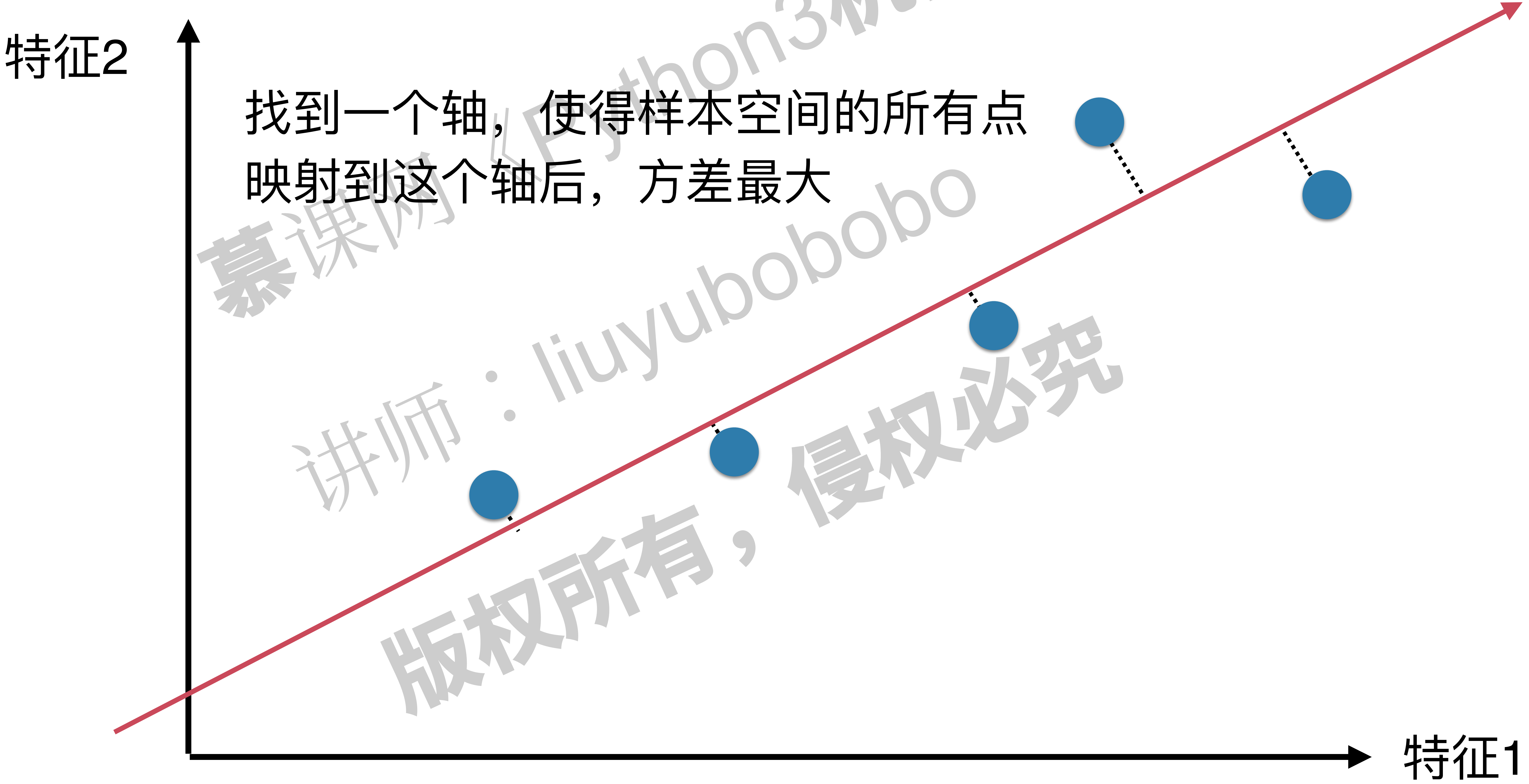
目标：求w，使得 $Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}(X^{(i)} \cdot w)^2$ 最大

$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}(X_1^{(i)}w_1 + X_2^{(i)}w_2 + \ldots + X_n^{(i)}w_n)^2$$

$$Var(X_{project}) = \frac{1}{m}\sum_{i=1}^{m}(\sum_{j=1}^{n}X_j^{(i)}w_j)^2$$

一个目标函数的最优化问题，使用梯度上升法解决

主成分分析



线性回归

梯度上升法解决主成分分析问题

主成分分析

目标：求w，使得 $f(X)=\frac{1}{m}\sum_{i=1}^{m}(X_1^{(i)}w_1+X_2^{(i)}w_2+\ldots+X_n^{(i)}w_n)^2$ 最大

$$\nabla f=\begin{pmatrix}\frac{\partial f}{\partial w_1}\\ \frac{\partial f}{\partial w_2}\\ \ldots\\ \frac{\partial f}{\partial w_n}\end{pmatrix}=\frac{2}{m}\begin{pmatrix}\sum_{i=1}^{m}(X_1^{(i)}w_1+X_2^{(i)}w_2+\ldots+X_n^{(i)}w_n)X_1^{(i)}\\ \sum_{i=1}^{m}(X_1^{(i)}w_1+X_2^{(i)}w_2+\ldots+X_n^{(i)}w_n)X_2^{(i)}\\ \ldots\\ \sum_{i=1}^{m}(X_1^{(i)}w_1+X_2^{(i)}w_2+\ldots+X_n^{(i)}w_n)X_n^{(i)}\end{pmatrix}$$

主成分分析

目标：求w，使得 $f(X)=\frac{1}{m}\sum_{i=1}^{m}(X_1^{(i)}w_1+X_2^{(i)}w_2+\ldots+X_n^{(i)}w_n)^2$ 最大

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \ldots \\ \frac{\partial f}{\partial w_n} \end{pmatrix} = \frac{2}{m}\begin{pmatrix} \sum_{i=1}^{m}(X^{(i)}w)X_1^{(i)} \\ \sum_{i=1}^{m}(X^{(i)}w)X_2^{(i)} \\ \ldots \\ \sum_{i=1}^{m}(X^{(i)}w)X_n^{(i)} \end{pmatrix}$$

主成分分析

$$= \frac{2}{m}\begin{pmatrix} \sum_{i=1}^{m}(X^{(i)}w)X_1^{(i)} \\ \sum_{i=1}^{m}(X^{(i)}w)X_2^{(i)} \\ \dots \\ \sum_{i=1}^{m}(X^{(i)}w)X_n^{(i)} \end{pmatrix}$$

$$\frac{2}{m}\cdot(\ X^{(1)}w,\quad X^{(2)}w,\quad X^{(3)}w,\quad \dots\quad X^{(m)}w\ )$$

$$\cdot\begin{pmatrix} X_1^{(1)} & X_2^{(1)} & X_3^{(1)} & \dots & X_n^{(1)} \\ X_1^{(2)} & X_2^{(2)} & X_3^{(2)} & \dots & X_n^{(2)} \\ X_1^{(3)} & X_2^{(3)} & X_3^{(3)} & \dots & X_n^{(3)} \\ \dots & \dots & \dots & \dots & \dots \\ X_1^{(m)} & X_2^{(m)} & X_3^{(m)} & \dots & X_n^{(m)} \end{pmatrix} = \frac{2}{m}\cdot(Xw)^T\cdot X$$

$$\frac{2}{m}\cdot X^T(Xw)$$

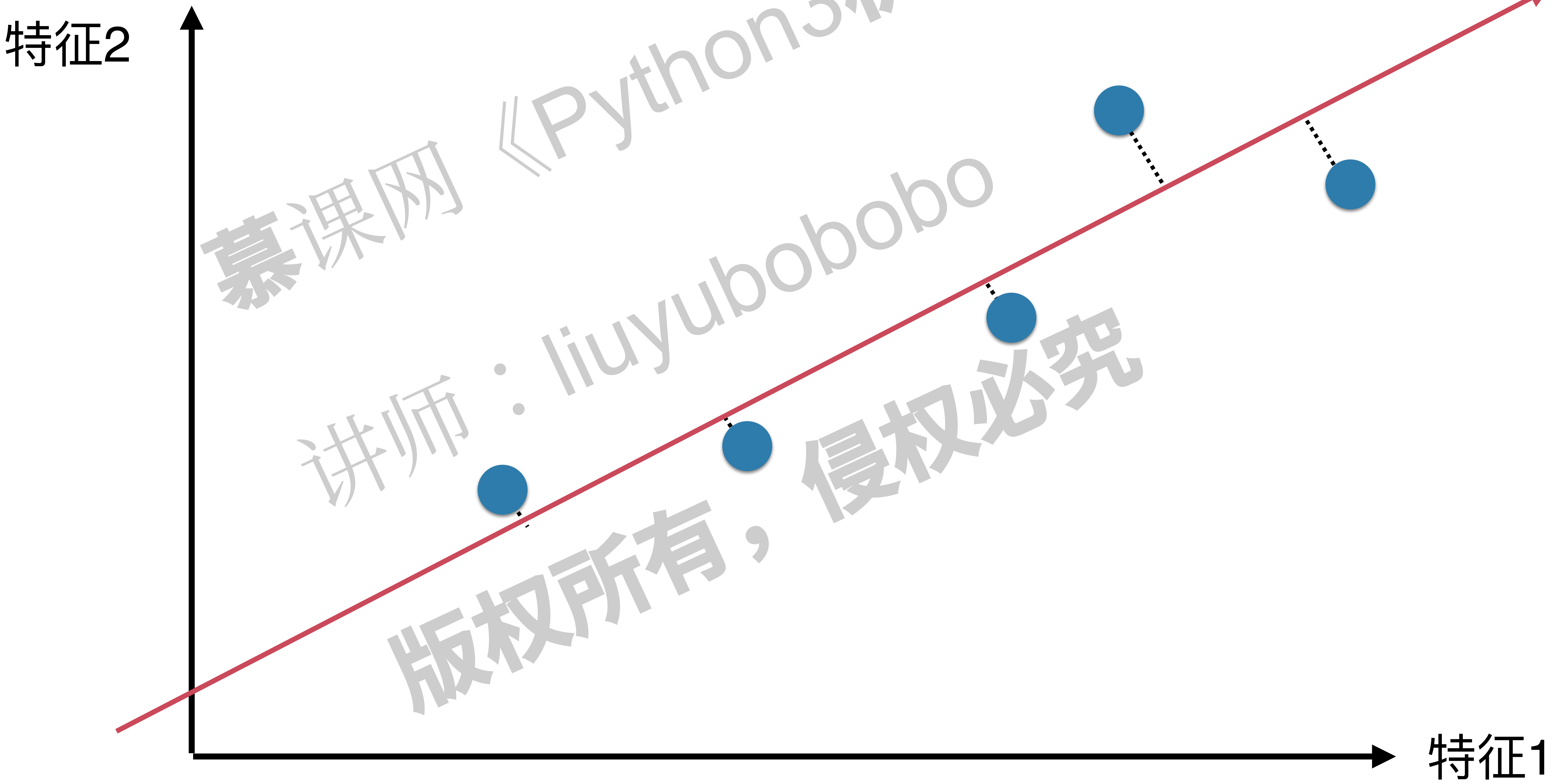
主成分分析

$$\nabla f = \frac{2}{m}\begin{pmatrix} \sum_{i=1}^{m}(X^{(i)}w)X_1^{(i)} \\ \sum_{i=1}^{m}(X^{(i)}w)X_2^{(i)} \\ \dots \\ \sum_{i=1}^{m}(X^{(i)}w)X_n^{(i)} \end{pmatrix} = \frac{2}{m}\cdot X^T(Xw)$$

实践：基于BGA实现PCA

求数据的前n个主成分

主成分分析



主成分分析

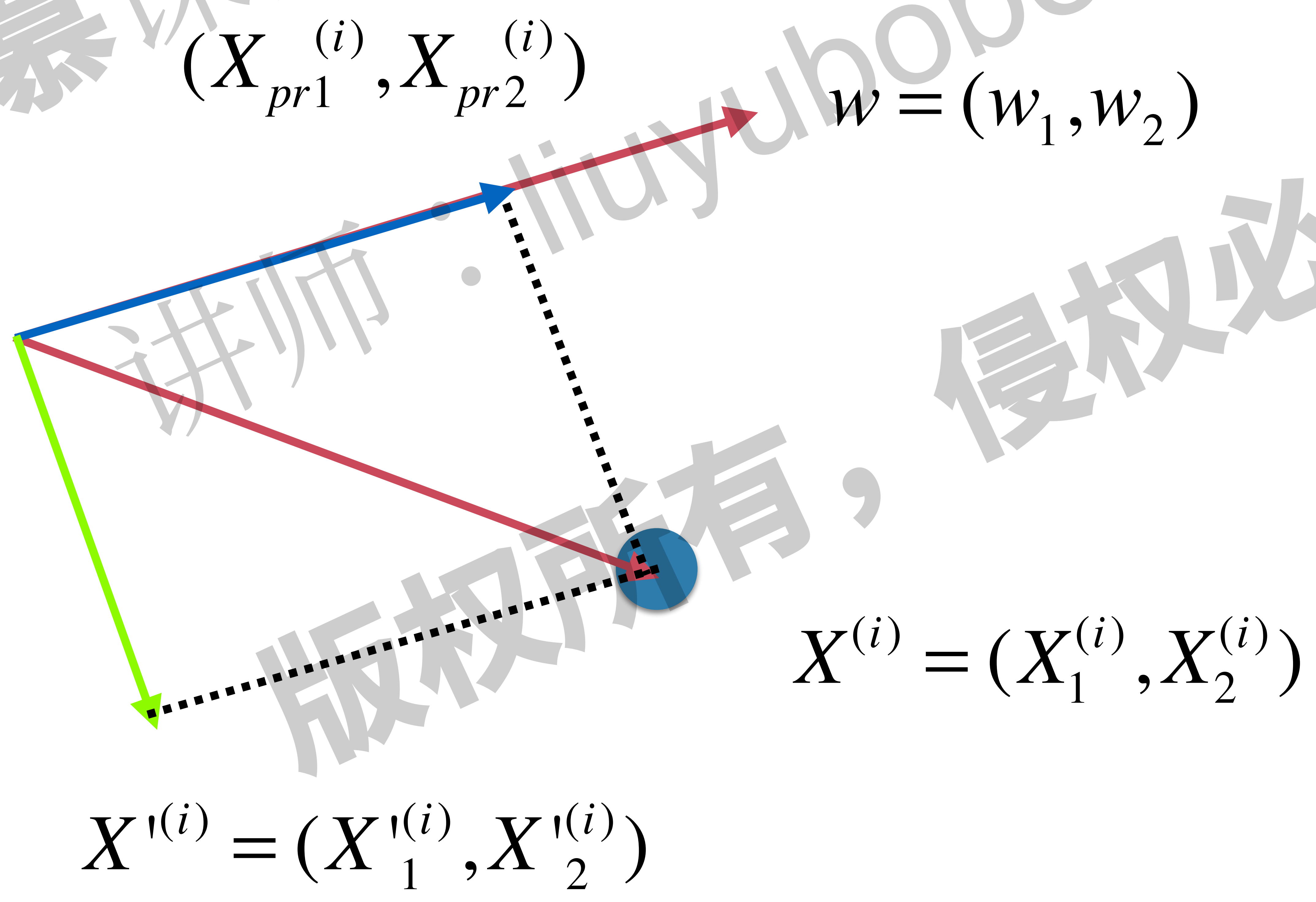
主成分分析

求出第一主成分以后，如何求出下一个主成分?

数据进行改变，将数据在第一个主成分上的分量去掉

主成分分析

数据进行改变，将数据在第一个主成分上的分量去掉



$$X^{(i)} \cdot w = \|X_{project}^{(i)}\|$$

$$X_{project}^{(i)} = \|X_{project}^{(i)}\| \cdot w$$

$$X'^{(i)} = X^{(i)} - X_{project}^{(i)}$$

主成分分析

求出第一主成分以后，如何求出下一个主成分?

数据进行改变，将数据在第一个主成分上的分量去掉

在新的数据上求第一主成分

实践：求前n个主成分

高维数据向低维数据映射

高维数据向低维数据映射

$$X = \begin{pmatrix} X_1^{(1)} & X_2^{(1)} & \dots & X_n^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \dots & X_n^{(2)} \\ \dots & \dots & \dots & \dots \\ X_1^{(m)} & X_2^{(m)} & \dots & X_n^{(m)} \end{pmatrix}$$

$$W_k = \begin{pmatrix} W_1^{(1)} & W_2^{(1)} & \dots & W_n^{(1)} \\ W_1^{(2)} & W_2^{(2)} & \dots & W_n^{(2)} \\ \dots & \dots & \dots & \dots \\ W_1^{(k)} & X_2^{(k)} & \dots & X_n^{(k)} \end{pmatrix}$$

$$X \cdot W_k^T = X_k$$

m*n　n*k　m*k

高维数据向低维数据映射

$$X_k = \begin{pmatrix} X_1^{(1)} & X_2^{(1)} & \dots & X_k^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \dots & X_k^{(2)} \\ \dots & \dots & \dots & \dots \\ X_1^{(m)} & X_2^{(m)} & \dots & X_k^{(m)} \end{pmatrix} \qquad W_k = \begin{pmatrix} W_1^{(1)} & W_2^{(1)} & \dots & W_n^{(1)} \\ W_1^{(2)} & W_2^{(2)} & \dots & W_n^{(2)} \\ \dots & \dots & \dots & \dots \\ W_1^{(k)} & X_2^{(k)} & \dots & X_n^{(k)} \end{pmatrix}$$

$$X_k \cdot W_k = X_m$$

m*k　　k*n　　　m*n

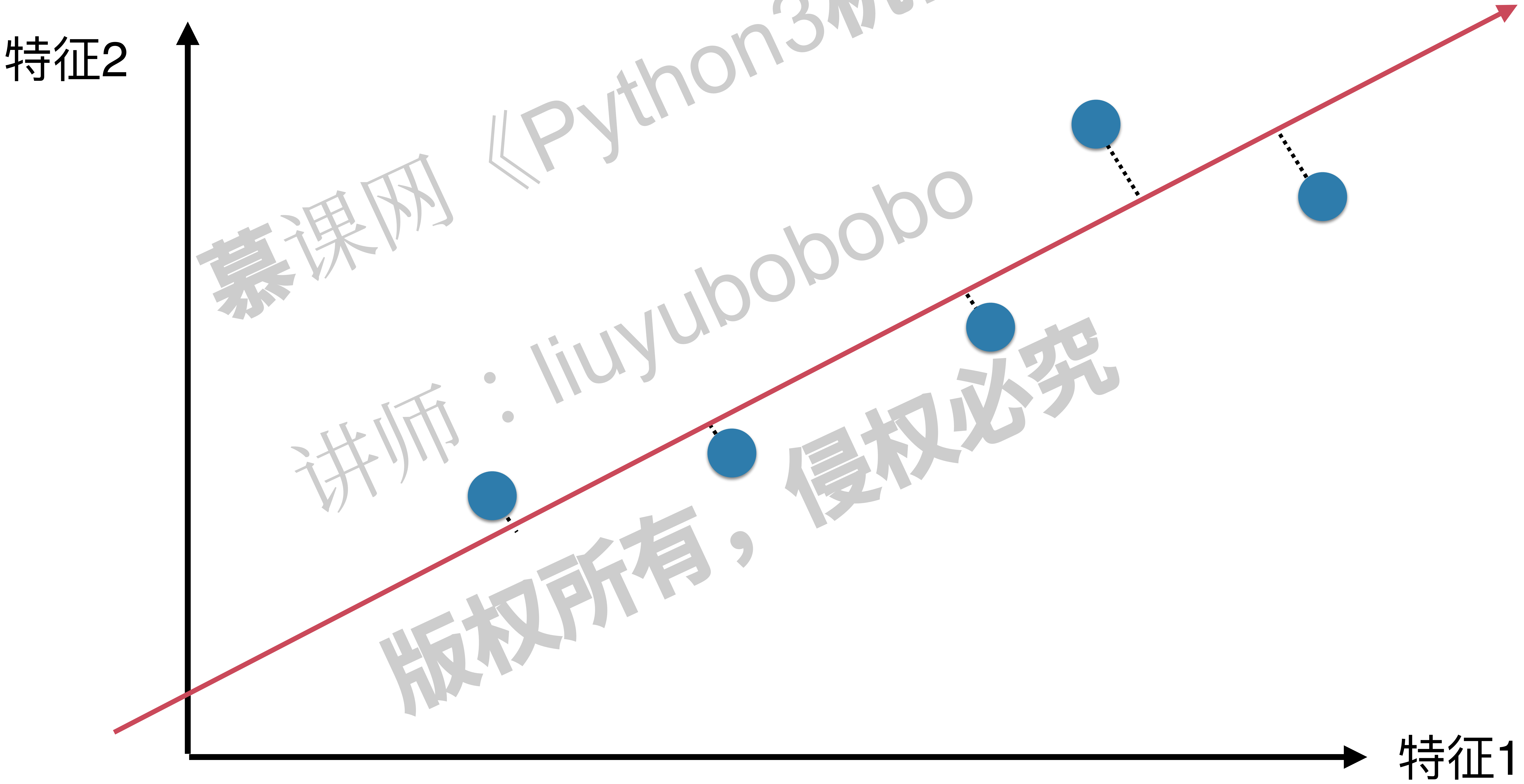
实践：封装PCA

scikit-learn中的PCA

MNIST数据集

使用PCA进行去噪

主成分分析



主成分分析

实践：使用PCA进行去噪

特征脸

高维数据向低维数据映射

$$X = \begin{pmatrix} X_1^{(1)} & X_2^{(1)} & \dots & X_n^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \dots & X_n^{(2)} \\ \dots & \dots & \dots & \dots \\ X_1^{(m)} & X_2^{(m)} & \dots & X_n^{(m)} \end{pmatrix} \qquad W_k = \begin{pmatrix} W_1^{(1)} & W_2^{(1)} & \dots & W_n^{(1)} \\ W_1^{(2)} & W_2^{(2)} & \dots & W_n^{(2)} \\ \dots & \dots & \dots & \dots \\ W_1^{(k)} & X_2^{(k)} & \dots & X_n^{(k)} \end{pmatrix}$$

$$X \cdot W_k^T = X_k$$

其他

欢迎大家关注我的个人公众号：是不是很酷

Python 3 玩儿转机器学习

liuyubobobo