

環境情報論第12回

応用編：主成分分析2

神山 翼, @t_kohyama,
tsubasa@is.ocha.ac.jp,

理3-703

今日も、データの重要な部分を
客観的に抜き出す方法を勉強します

応用編：主成分分析2

新しい x' 軸に射影することで x' 成分をもとめる

主成分分析の性質

二重直交性 + 分散の和の保存

データの散らばりを記述するために

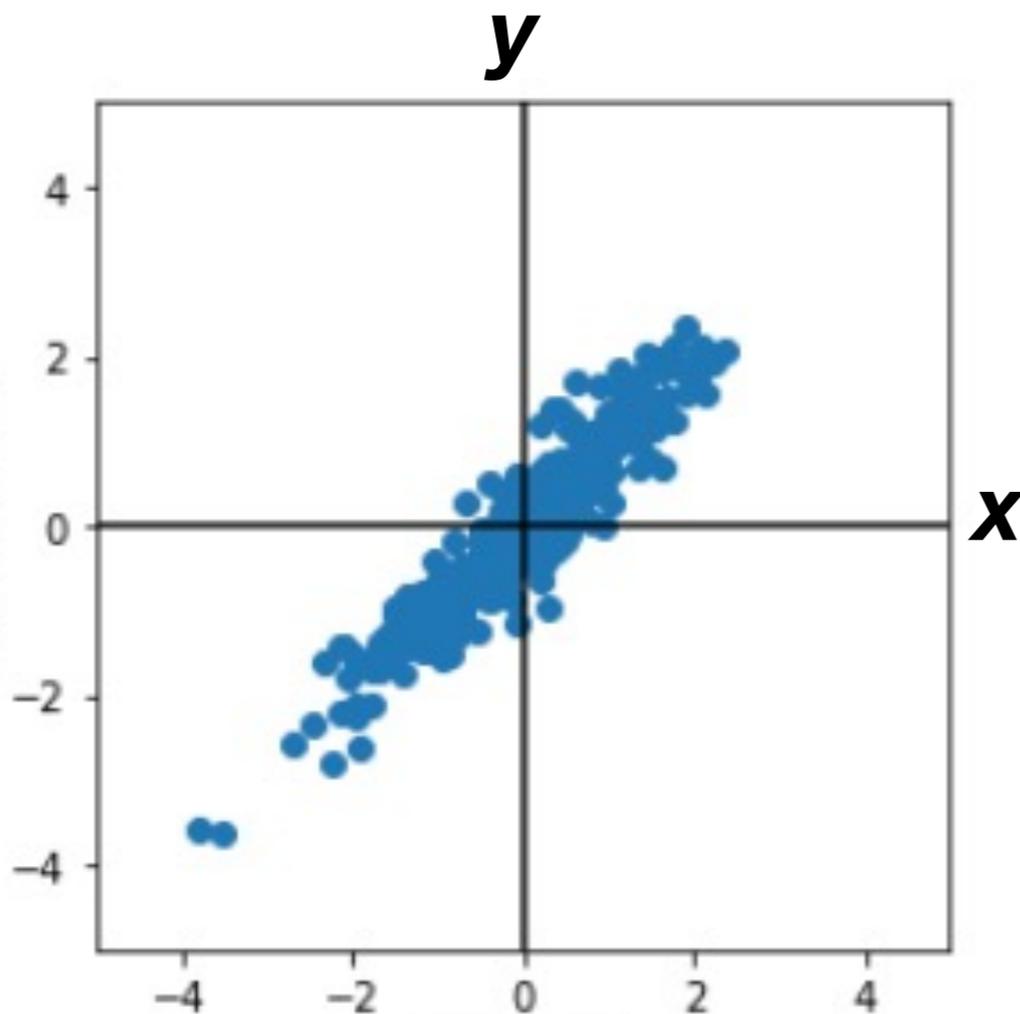
空間内の座標変換を行い

基底を自由に取り直せるようになるのが目標

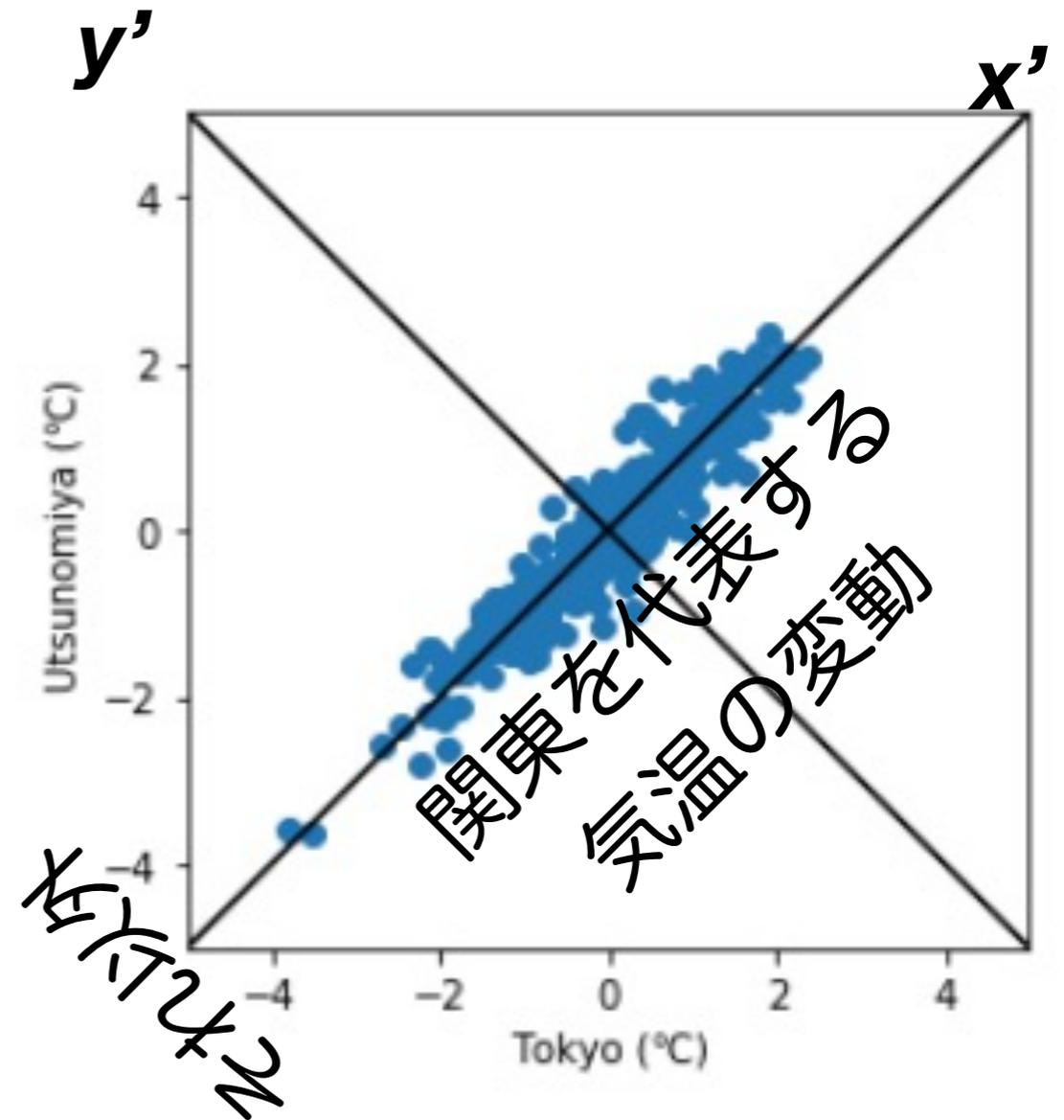
主成分分析の流れ（前回の復習）

座標軸を最大分散方向とその直交方向
になるように回転させる分析

宇都宮の気温偏差

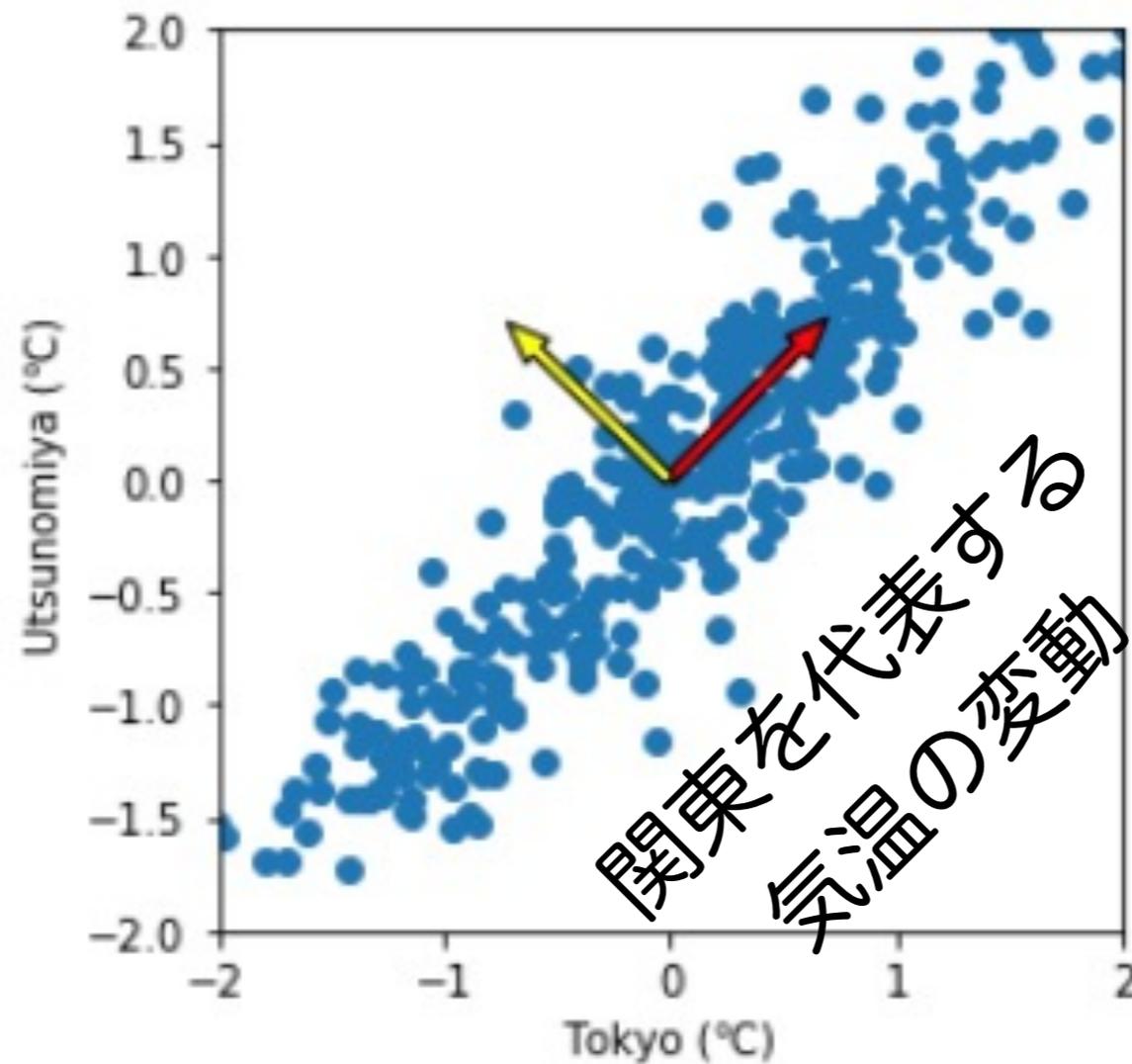


東京の気温偏差



主成分分析の流れ（前回の復習）

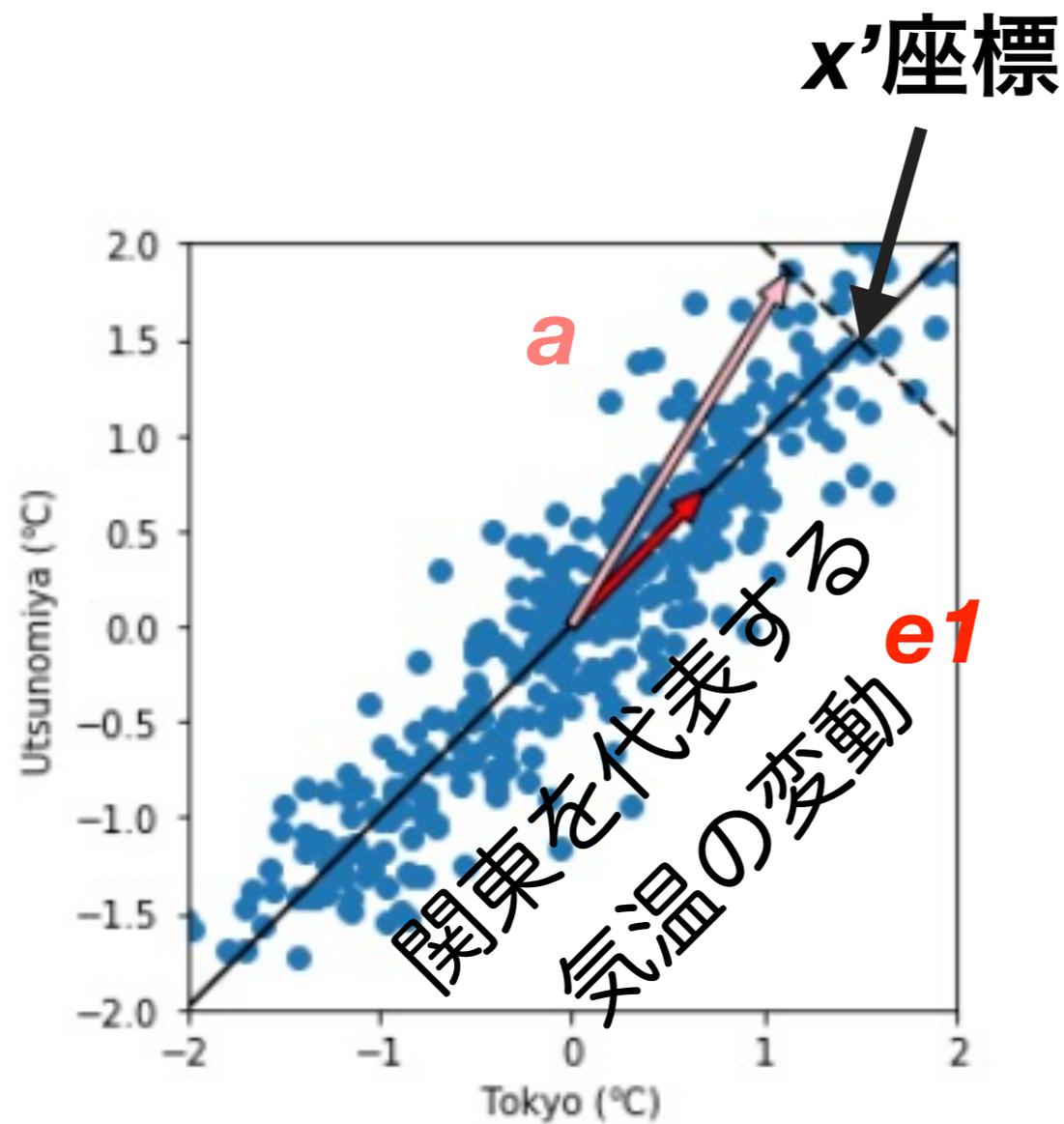
共分散行列の固有値固有ベクトルを計算し、
分散最大方向を抽出



分散最大方向の単位ベクトル e_1 （赤）と
その直交方向の単位ベクトル e_2 （黄色）

データ点 a の x' 座標を求める方法

基底ベクトルとの内積を取れば良い

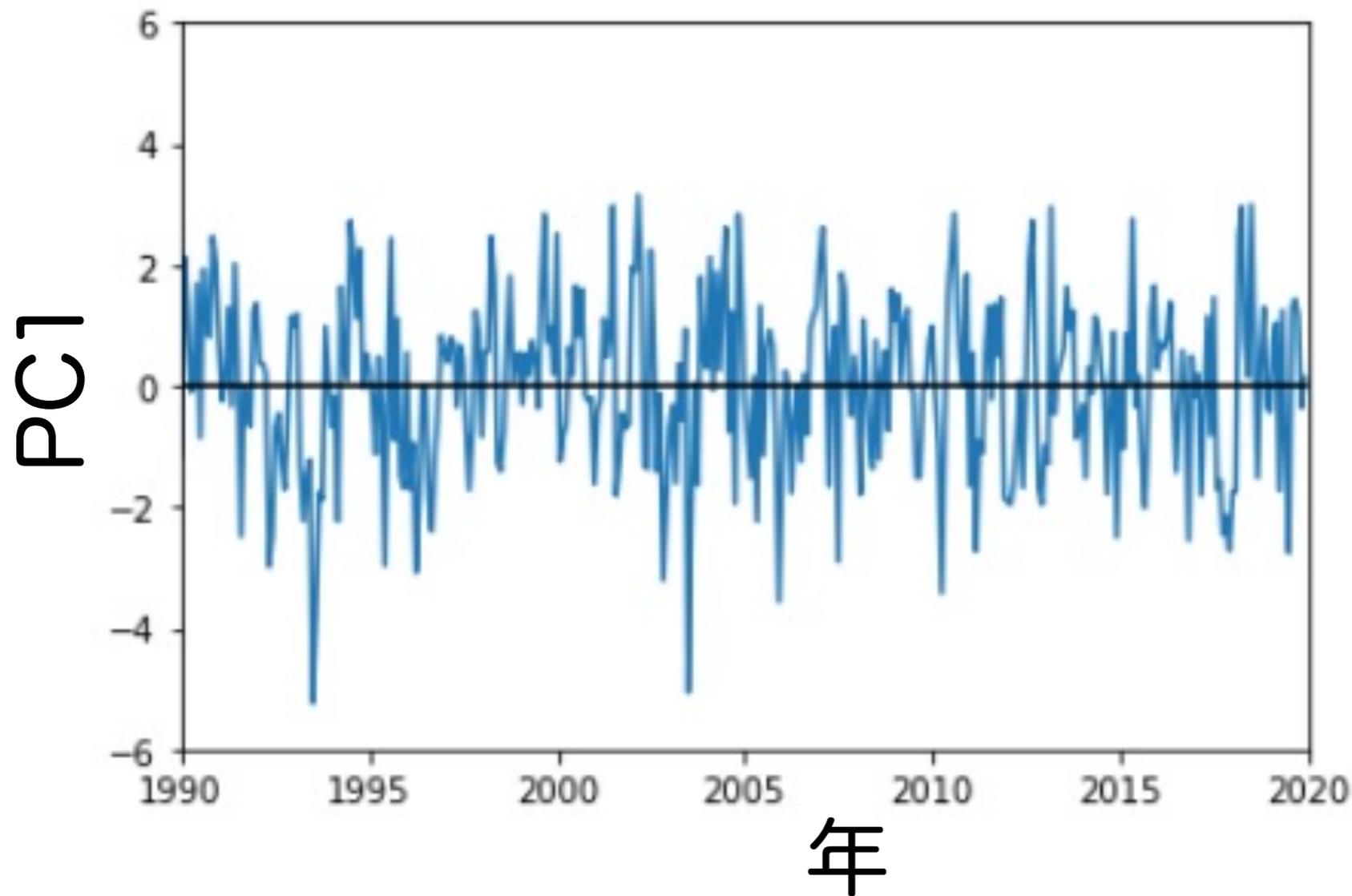


$$x' = |\vec{a}| \cos \theta = \vec{e}_1 \cdot \vec{a}$$

これを基底 e_1 に
射影するという

第一主成分の時系列 = PC1

全ての月について、x'座標を計算して並べた時系列



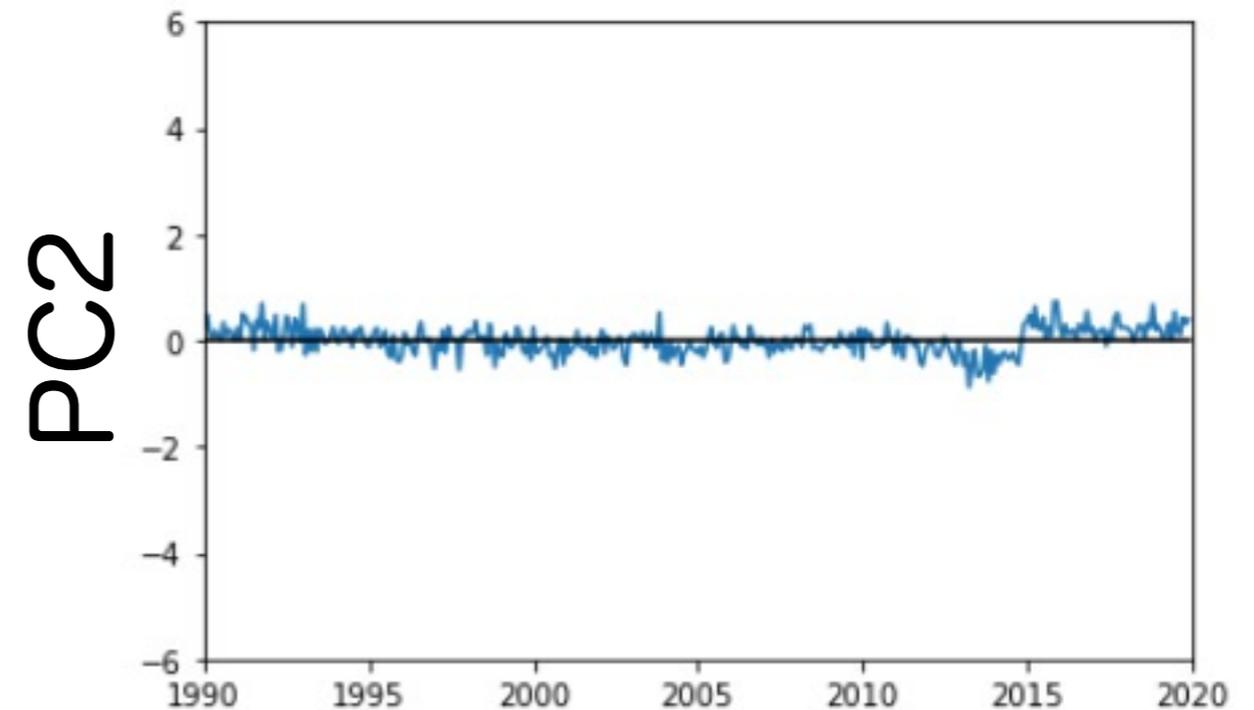
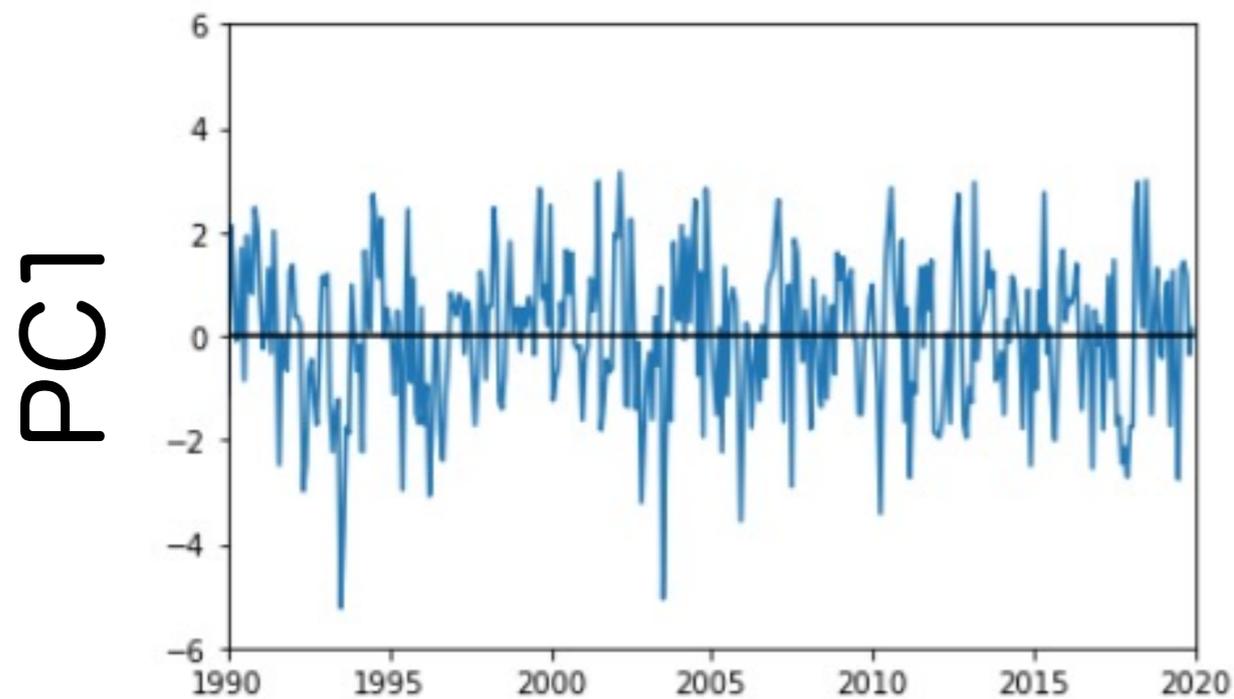
「関東代表」の
時系列

PC1が正の時は
東京も宇都宮も
暖かい

※ PC = Principal Component (主成分)

PC1とPC2

PC2は、PC1に大きな分散を押し付けた後の「出がらし」



$$X = \vec{e}_1 PC1 + \vec{e}_2 PC2$$

元の気温偏差時系列を、第1項の「東京と宇都宮が同位相で変化する第1主成分」（振幅が大きい）、第2項の「東京と宇都宮が逆位相で変化する第2主成分」（振幅が小さい）に分解できた

主成分分析の性質1: 二重直交性

固有ベクトル同士・PC時系列同士は
それぞれ直交する

```
np.dot(e1, e2) # 固有ベクトル同士の内積はゼロ (前回のC問題)
```

```
0.0
```

```
np.dot(pc1, pc2) # PC時系列同士の内積もゼロ
```

```
-7.149836278586008e-14
```

すなわち, PC時系列同士は無相関
**分散の大きい順に, 互いに相関のない成分に
分解するのが主成分分析**

主成分分析の性質2:

固有値は対応する固有ベクトル方向の分散

行列 Λ には, 対角成分に各PCの分散が入っている

```
print(np.var(pc1))  
print(np.var(pc2))
```

```
2.0513924657032288  
0.06501294064564263
```

```
print(L)
```

```
[[2.05139247  0.          ]  
 [0.          0.06501294]]
```

主成分分析の性質3: 分散の和が保存される

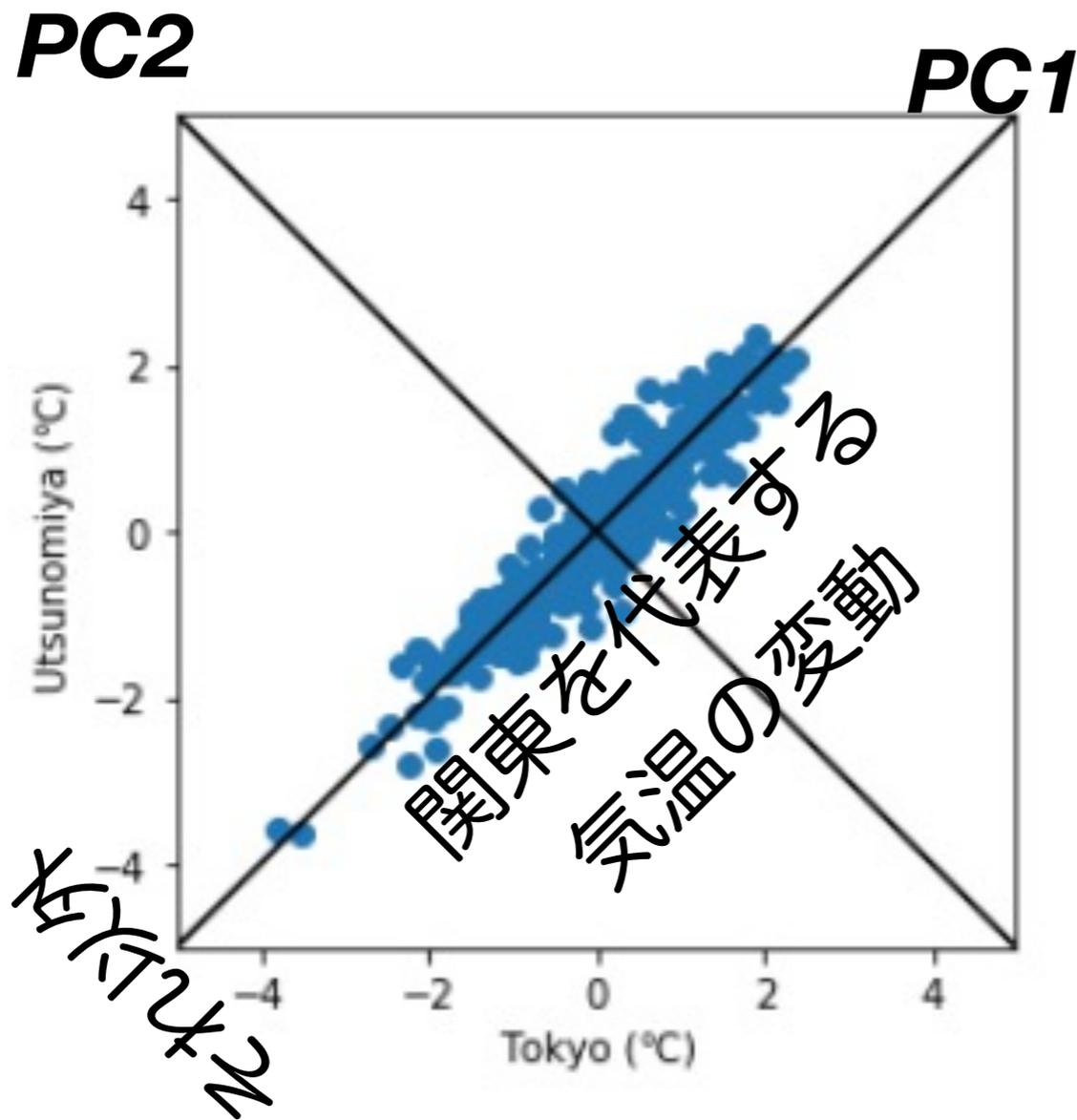
東京の気温偏差の分散 + 宇都宮の気温偏差の分散
= PC1の分散 + PC2の分散

$$EV_i := \frac{\text{var}(\vec{PC}_i)}{\sum_{j=1}^M \text{var}(\vec{x}_j)} = \frac{\lambda_i}{\sum_{j=1}^M \lambda_j}$$

分散の和に対する各PCの分散の比のことを
寄与率 (explained variance) という。

今回の例だと，PC1で
データの散らばりをほとんど全て説明できてしまう

PC1の寄与率は97%



本来2つだったはずの
時系列の情報が
1つの時系列でほとんど
説明できてしまう
→次元削減成功！

発展：2次元気象場における主成分分析

インデックスの定義などに頻繁に用いられる

データ行列

時刻

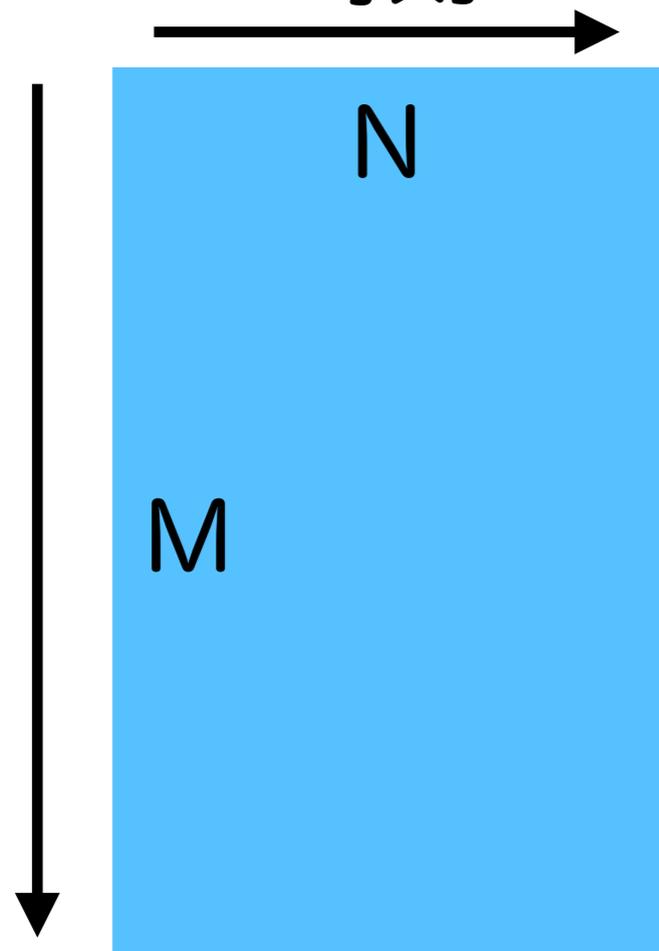
(授業資料)

(B問題)

(2次元気象場)

$X =$

場所



東京

宇都宮

(M=2)

札幌

仙台

東京

大阪

福岡

那覇

(M=6)

90°N, 1°E

90°N, 2°E

90°N, 3°E

⋮

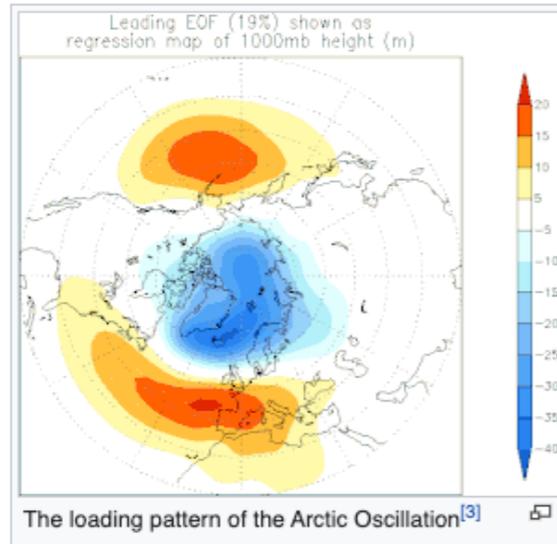
90°S, 1°W

(M=360×180)

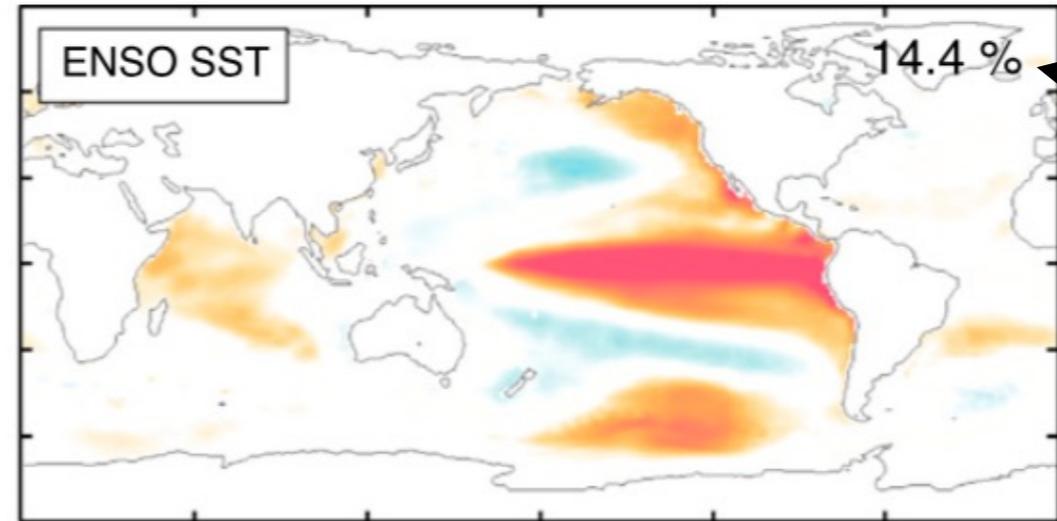
※共分散行列を $\cos \phi$ で重み付けする必要がある (ϕ は緯度)

発展：2次元気象場における主成分分析

固有ベクトル

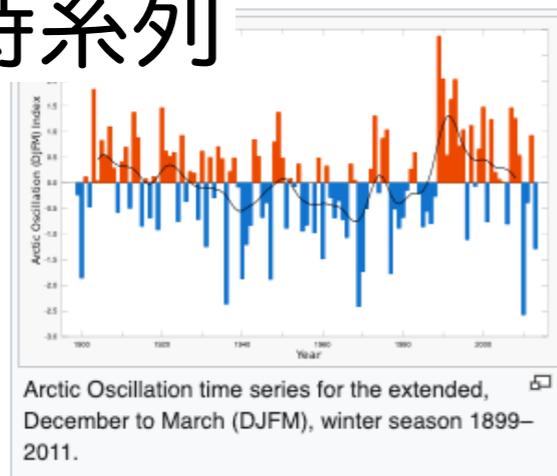


固有ベクトル



寄与率

PC1時系列



**世界の海面水温の第一主成分
=エルニーニョ南方振動のインデックス**

北半球における

地表気圧の第一主成分

= 北極振動のインデックス

さらに詳しく知りたい方は
大学院で... (笑)

たのしい課題たち

A: PC時系列と寄与率の説明+講義の感想

B: 6次元データにおける主成分の解釈

C: 次元削減の検証

D: 自由課題

今日も、データの重要な部分を
客観的に抜き出す方法を勉強します

応用編：主成分分析2

新しい x' 軸に射影することで x' 成分をもとめる

主成分分析の性質

二重直交性 + 分散の和の保存

データの散らばりを記述するために

空間内の座標変換を行い

基底を自由に取り直せるようになるのが目標

本日の導入パートは以上です。
何でも良いのでZoomの方に
授業に関係のあるコメントを
してください（出席代わり）。

コメント拾いが終わったら、
早速今日のプログラミングに進みましょう。