

自然言語処理 —準備、数学—

<https://yo-sato.com/>

数学準備の位置づけ

- 機械学習、ニューラルネットは、コードを書くのはほぼパッケージのクラスを利用すればいいのですが、やっていることを理解するには、高校数学の一部だけ思い出しておくと、納得感が得られます。
- 今後のトピックのスライドの予備知識のところに出てくるので、思い出してほしい最低限を以下に掲載し、さらにちゃんとした数学の先生が解説した動画を視聴することを演習課題とします。
- 自信がある方は、確認クイズを最初にやってください。

シグマ記号は合計の略記法

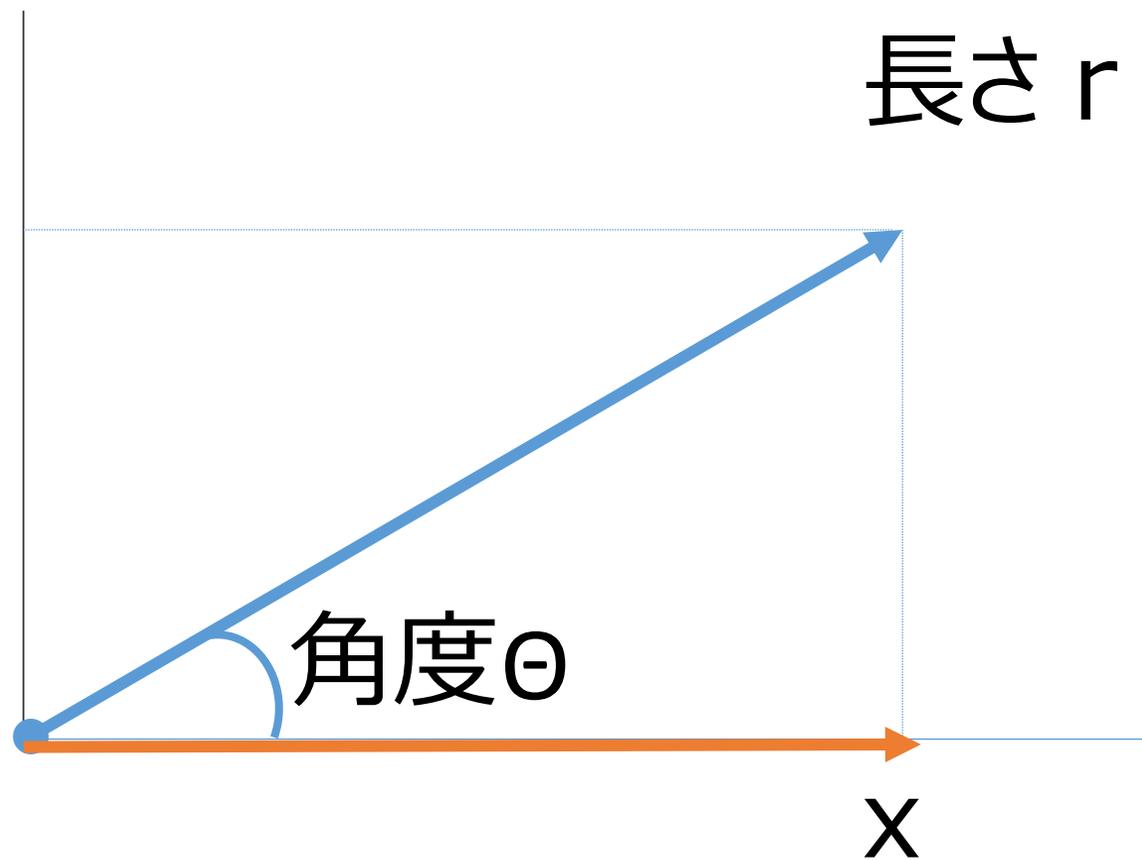
$$y = x_1 + x_2 + x_3$$

=>

$$y = \sum_{i=1}^3 x_i \quad \text{または} \quad y = \sum_i x_i$$

三角関数コサイン

$$\cos \theta = \frac{x}{r}$$



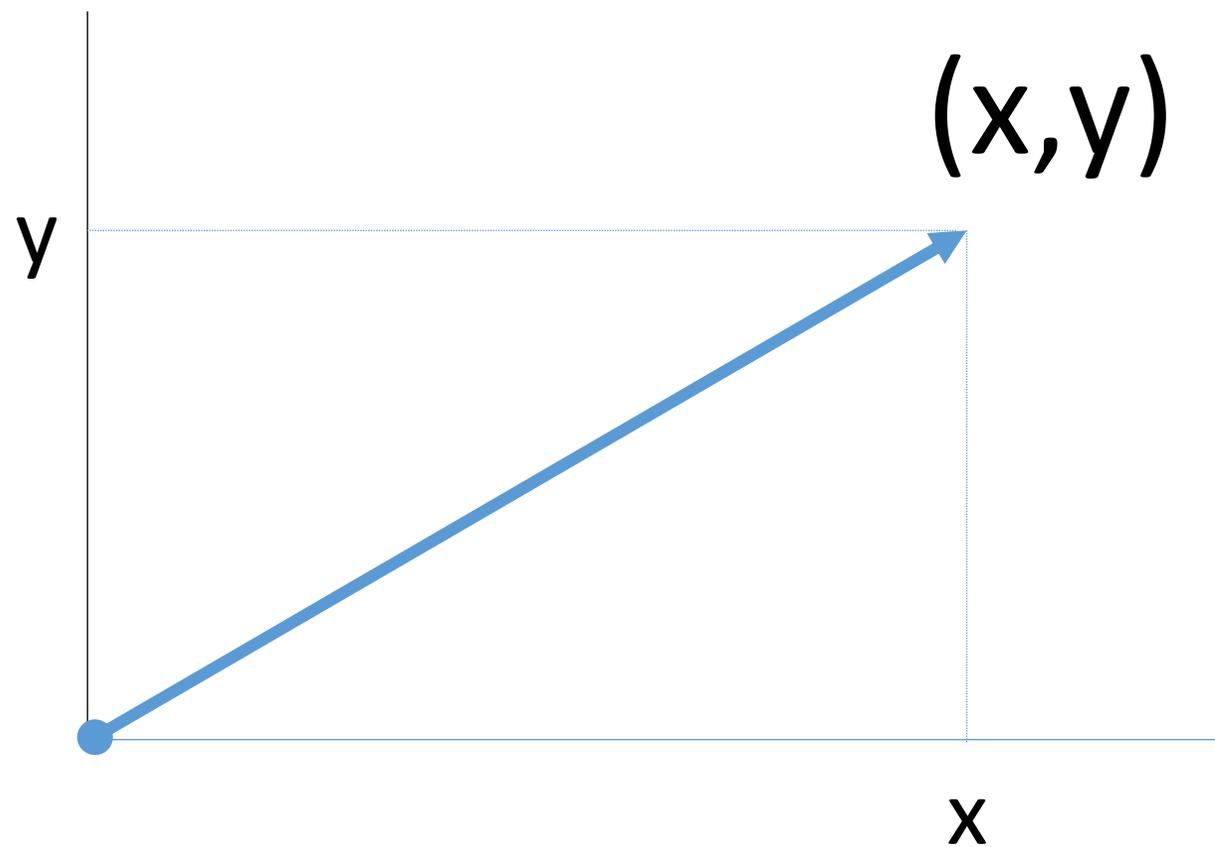
指数と対数

$$y = a^x \iff x = \log_a y$$

課題 1

- 以下を視聴してください。 <https://edupa.org/> 数II
 - 対数の定義 : <https://www.youtube.com/watch?v=NmFfjffB8lg> (16分)
 - 対数のグラフ : <https://www.youtube.com/watch?v=JFHR8l9bZfM> (12分)

N次元ベクトル



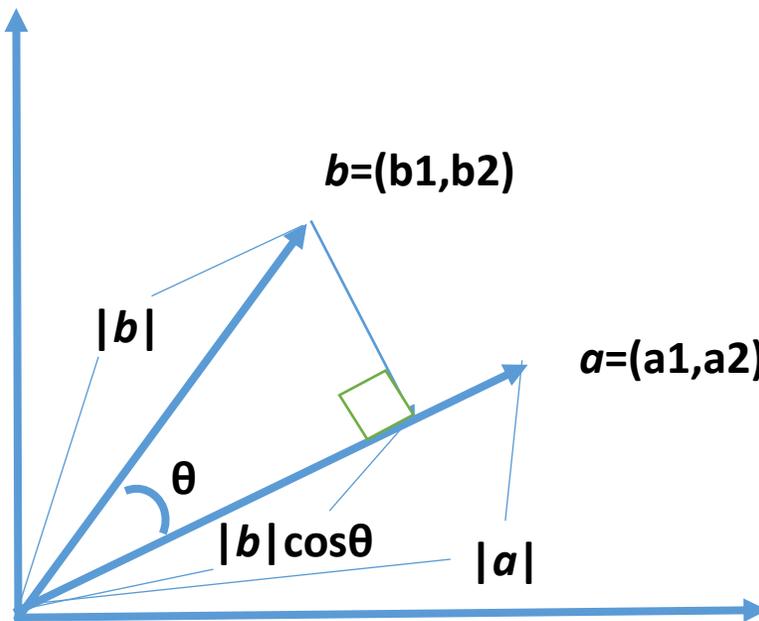
2次元の場合

課題 2

- 以下を視聴してください。 <https://edupa.org/> 数B
 - ベクトルの概念：
<https://www.youtube.com/watch?v=gqWklJtBIC0&feature=youtu.be> (18分)

ベクトルの内積：それぞれの個性の強さに相性を加味した量

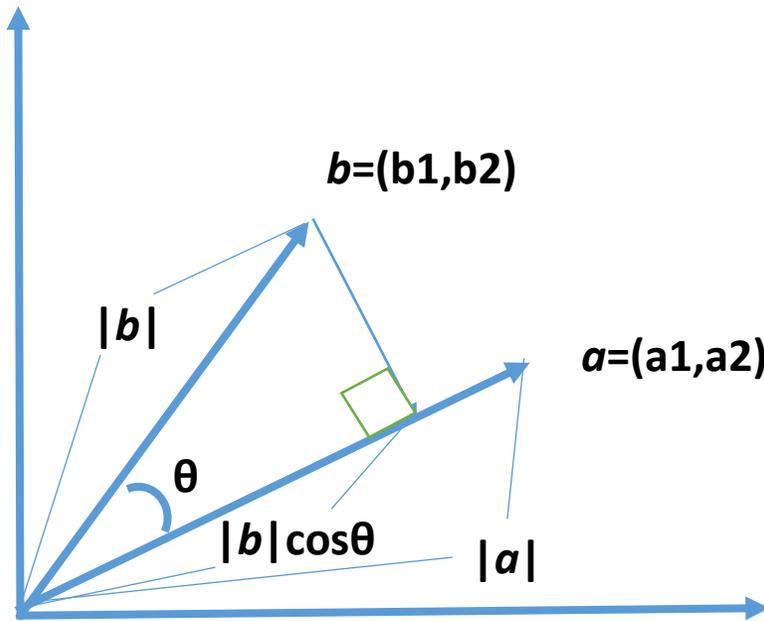
ベクトル $a = (a_1, a_2)$, $b = (b_1, b_2)$ の内積 $a \cdot b = a_1 b_1 + a_2 b_2 = |a| |b| \cos \theta$



相性がいい、同じ方向を向いているならば、 $\cos \theta = 1$ 。内積は $|a| |b|$ 。

相性が悪く、直角に向いているならば、 $\cos(\theta) = 0$ 。内積は0。

相性の評価尺度: コサイン類似度

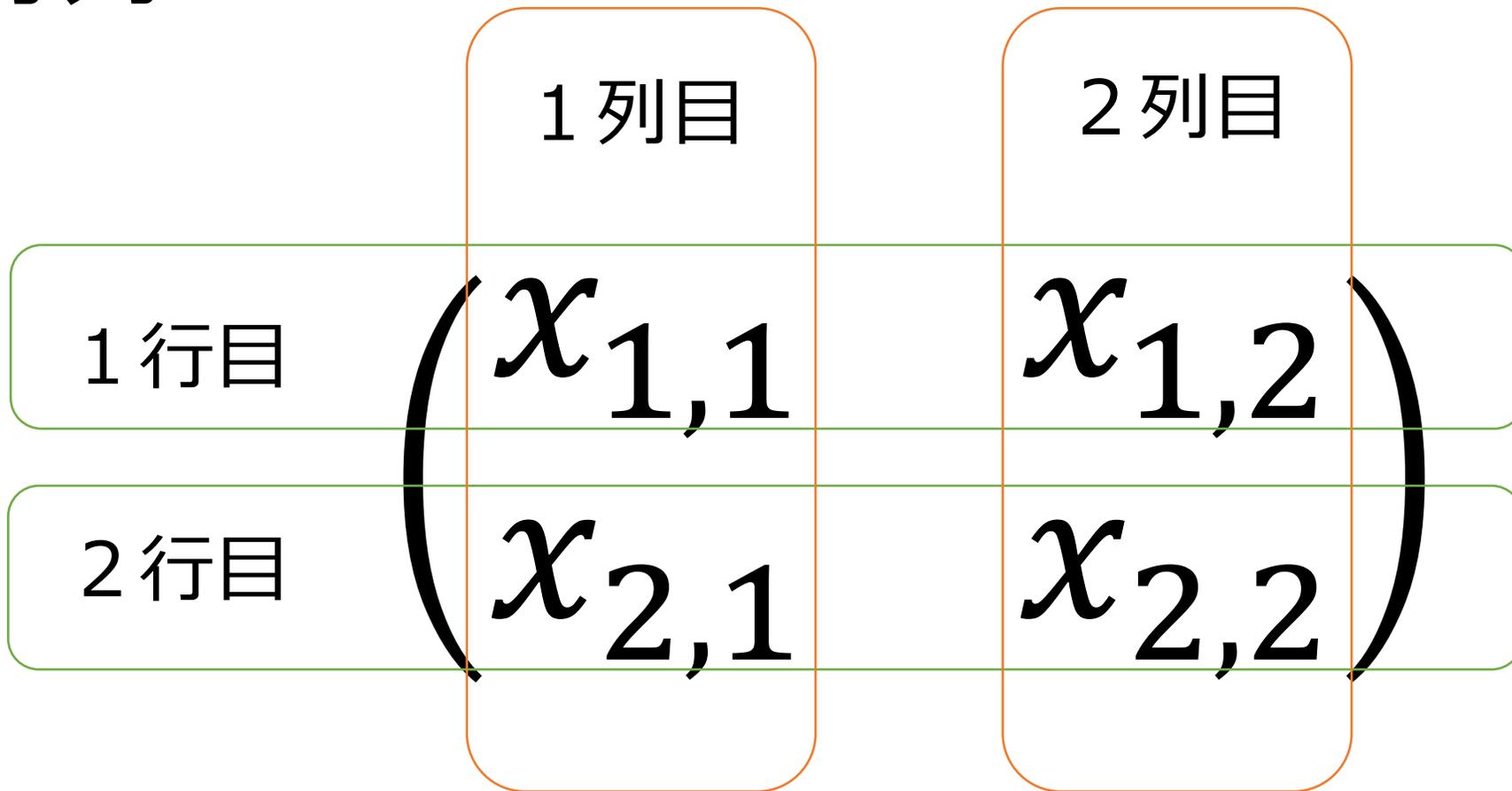


内積 $a \cdot b = a_1b_1 + a_2b_2 = |a||b|\cos\theta$

$$\text{CosineSimilarity}(a, b) = \frac{a \cdot b}{|a||b|}$$

ベクトル a と b の角度のCosineは、
相性が良く同じ方向を向いていれば 1
直角に背いていれば 0
相性が真逆ならば、 -1

行列



2x2の行列

課題 3

- 以下を視聴してください。 <https://edupa.org/> 数C
 - 行列の概念 : <https://www.youtube.com/watch?v=S7gqMHxIL6Q> (16分)
 - 行列の積 : <https://www.youtube.com/watch?v=a71xF5FNxYk> (18分)

線形性

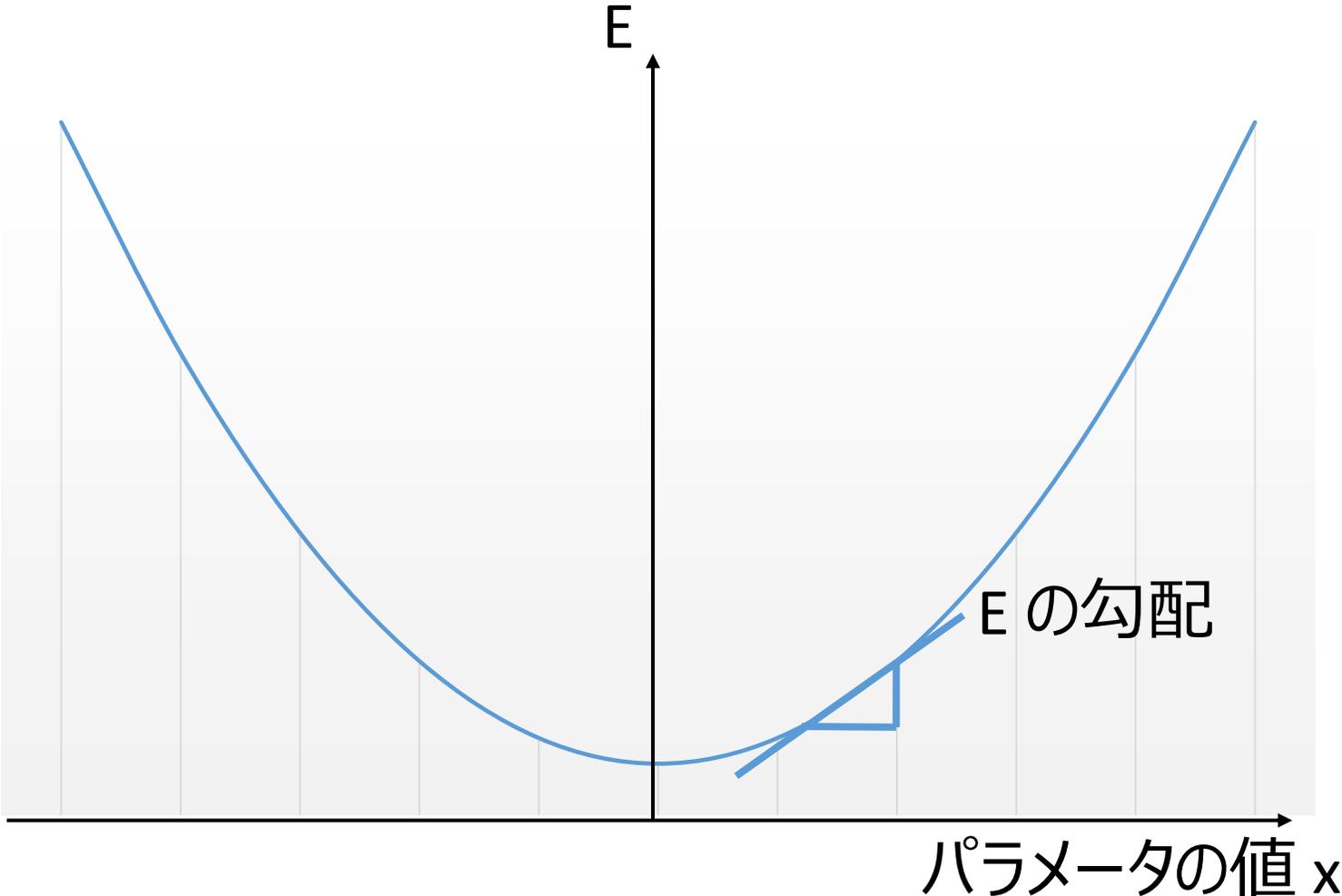
写像 f が、以下の性質を満たすとき、線形であるという。

$$f(x+y)=f(x)+f(y)$$

$$f(ax)=af(x)$$

参考： 高校数学における線形性の8つの例 <https://manabitimes.jp/math/684>

微分、傾き



Eの変分

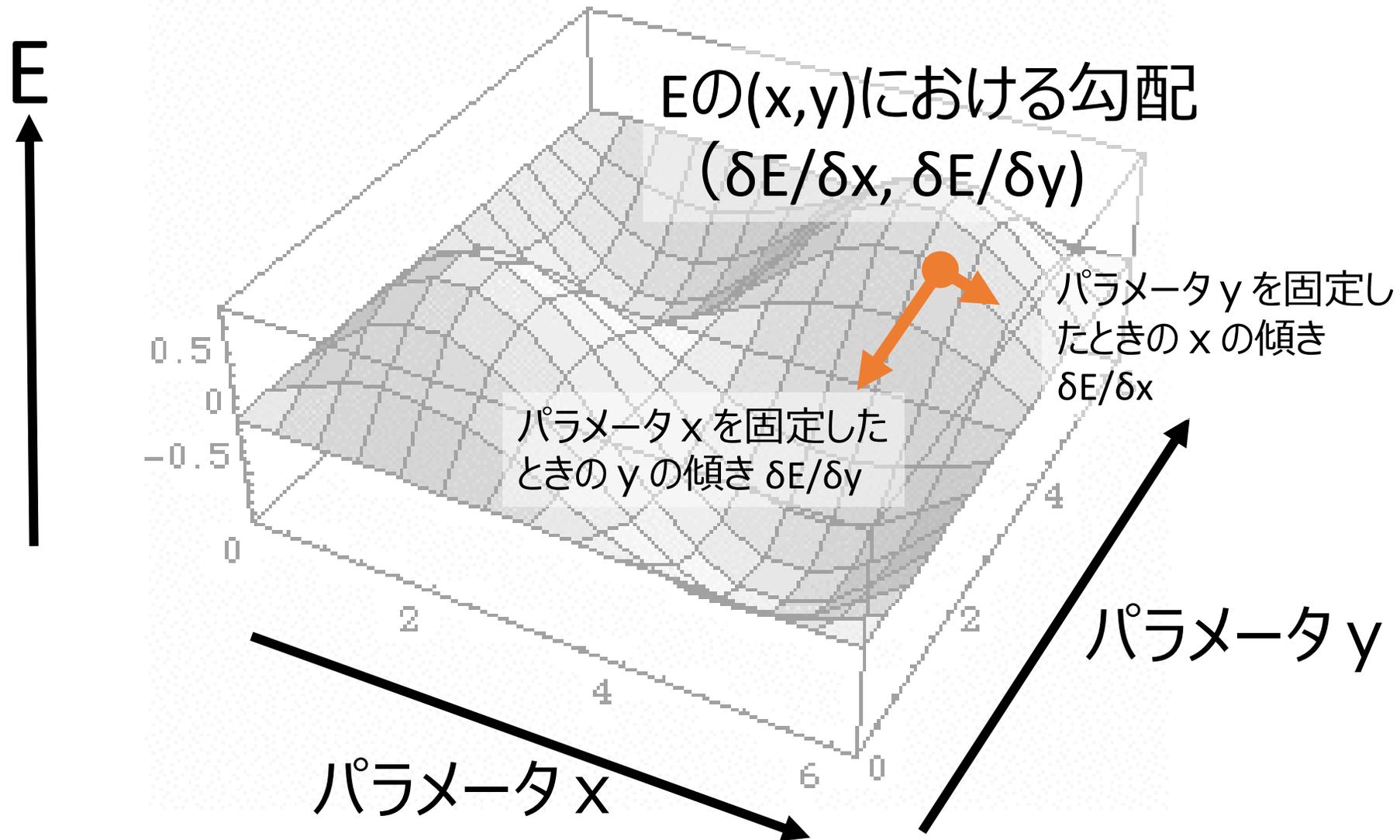
$$\frac{\Delta E}{\Delta x}$$

xの変分

課題 4

- 以下を視聴してください。 <https://edupa.org/> 数II
 - 微分の概念 : <https://www.youtube.com/watch?v=4sjS0nzd4SI> (14分)

偏微分、勾配



大学数学つまみ食い課題 5

- 以下を視聴してください。
 - 【大学数学】偏微分とは何か【解析学】
<https://www.youtube.com/watch?v=UWFTIElruyc> (18分)

AIプログラミングと数学の関連

- 【AI】なんで線形代数はプログラミングに大事？気になる機械学習、ディープラーニングとの関係性まで徹底解説！
https://www.geekly.co.jp/column/cat-technology/1902_049/
 - 線形代数といっても、教科書の全部は必要ありません。行列の積が、ニューラルネットの構成要素としての基本の全結合層の演算そのもの。
 - 微分積分といっても、教科書の全部は必要ありません。偏微分の考え方が、ニューラルネットの学習の逆伝播（バックプロパゲーション）アルゴリズムの基礎になっている。ただ、逆伝播は今時のクラスライブラリパッケージのメソッドの一機能になっているので、動かすのはメソッドを呼ぶだけ。何をやっているかの雰囲気をつかんでしまえば、内部の処理は忘れても大丈夫。
- 徹底的に理解したければ、<https://python.atelierkobato.com/> とか。

確認クイズ

- 確認クイズをやってください。