

[Movie](#) in the mp4 format.

動画: [Neural Network とは?](#) (youtube, 限定公開).

[プログラム](#) (for pytorch on python)

手書き数字の認識: [pytorch](#) で [MNIST](#) (動画で紹介しているページ)

0050.png

深層学習 (deep learning) $\subset$ ニューラルネットワークによる学習 (Neural net) $\rightarrow$ NN

における数理アルゴリズム (mathematical algorithm)

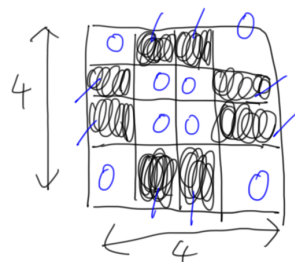
1. 最適化のアルゴリズム (入門)

2. NNによる学習の原理

キーワード: forward
backpropagation

3. backpropagation を知る

何かいっているのか? ティモ
画像という
手書き数字の認識



MNIST ティモ

@ tensorflow

○ ティモ

入力

0110 1001 1001 0110

↓ F

この入力に文字が対応

28x28 自己 0と1で表現

→ 0~255 = $2^8 - 1$

の値を表現

→ 黒

0100.png

Google 深層学習

ニューラルネットワーク -... エレコム USB Type-... 第3回 PyTorchによる... ライトニングpytorch入... 深層学習 - Google 検索 pytorch mnist - Goog... cifar10 - Google 検索

Google 深層学習

すべて ニュース 画像 ショッピング 動画 もっと見る 設定 ツール

約 4,660,000 件 (0.53 秒)

ディープラーニング(深層学習)とは、人間が自然に行うタスクをコンピュータに学習させる機械学習の手法のひとつです。人工知能(AI)の急速な発展を支える技術であり、その進歩により様々な分野への実用化が進んでいます。

jp.mathworks.com › discovery › deep-learning
[ディープラーニング - これだけは知っておきたい3つのこと ...](#)

強関スニペットについて フィードバック

ledge.ai › Topic Keywords › AI:人工知能

[ディープラーニング\(深層学習\)とは | 基本知識・仕組み ...](#)

2020/02/10 — ディープラーニング (Deep Learning) または 深層学習とは、ニューラルネットワークを多層に結合し表現・学習能力を高めた機械学習の手法です。歴史から仕組み、人工知能 (AI) 、ニューラルネットワークや機械学習との ...

ja.wikipedia.org › wiki › ディープラーニング
[ディープラーニング - Wikipedia](#)

ディープラーニング

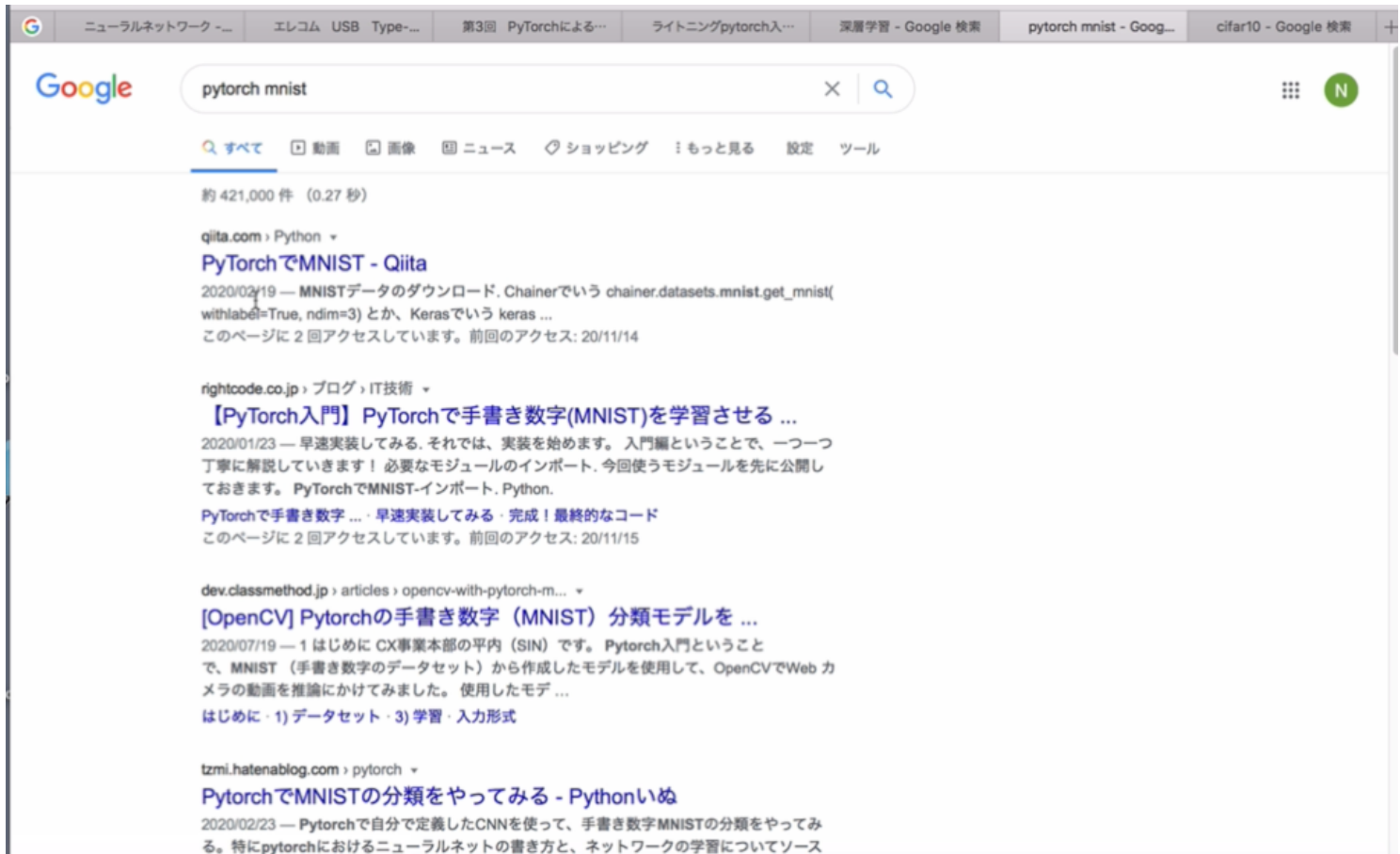
ディープラーニングまたは深層学習とは、多層の人工ニューラルネットワークによる機械学習手法である。要素技術としてはバックプロパゲーションなど、20世紀のうちに開発されていたものの、4層以上の深層ニューラルネットについて、局所最適解や勾配消失などの技術的な問題によって十分学習させられず、性能も芳しくなかった。 [ウィキペディア](#)

他の人はこちらも検索 他 15 件以上を表示

機械学習 人工知能 Python ロボット工学

フィードバック

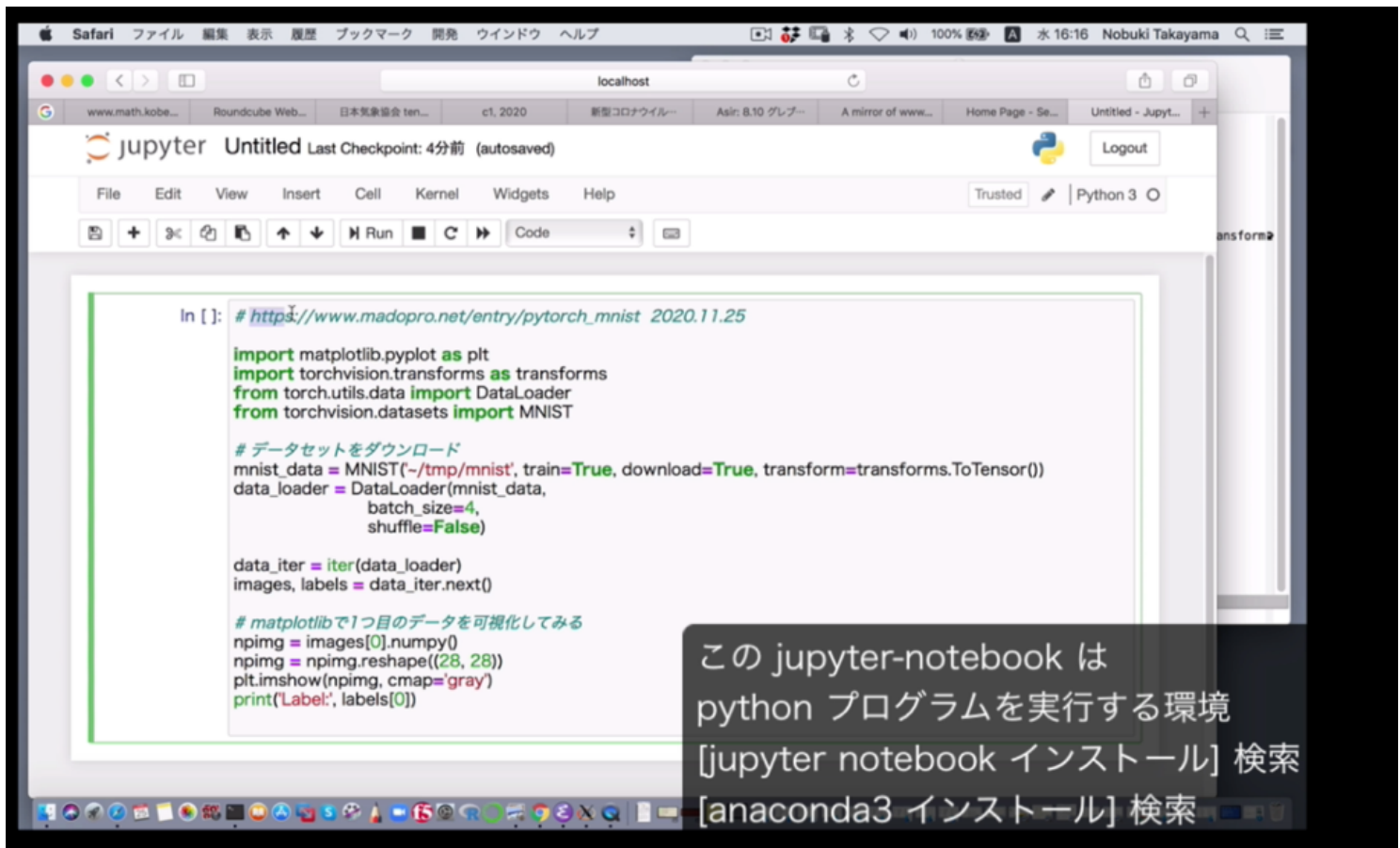
0200.png



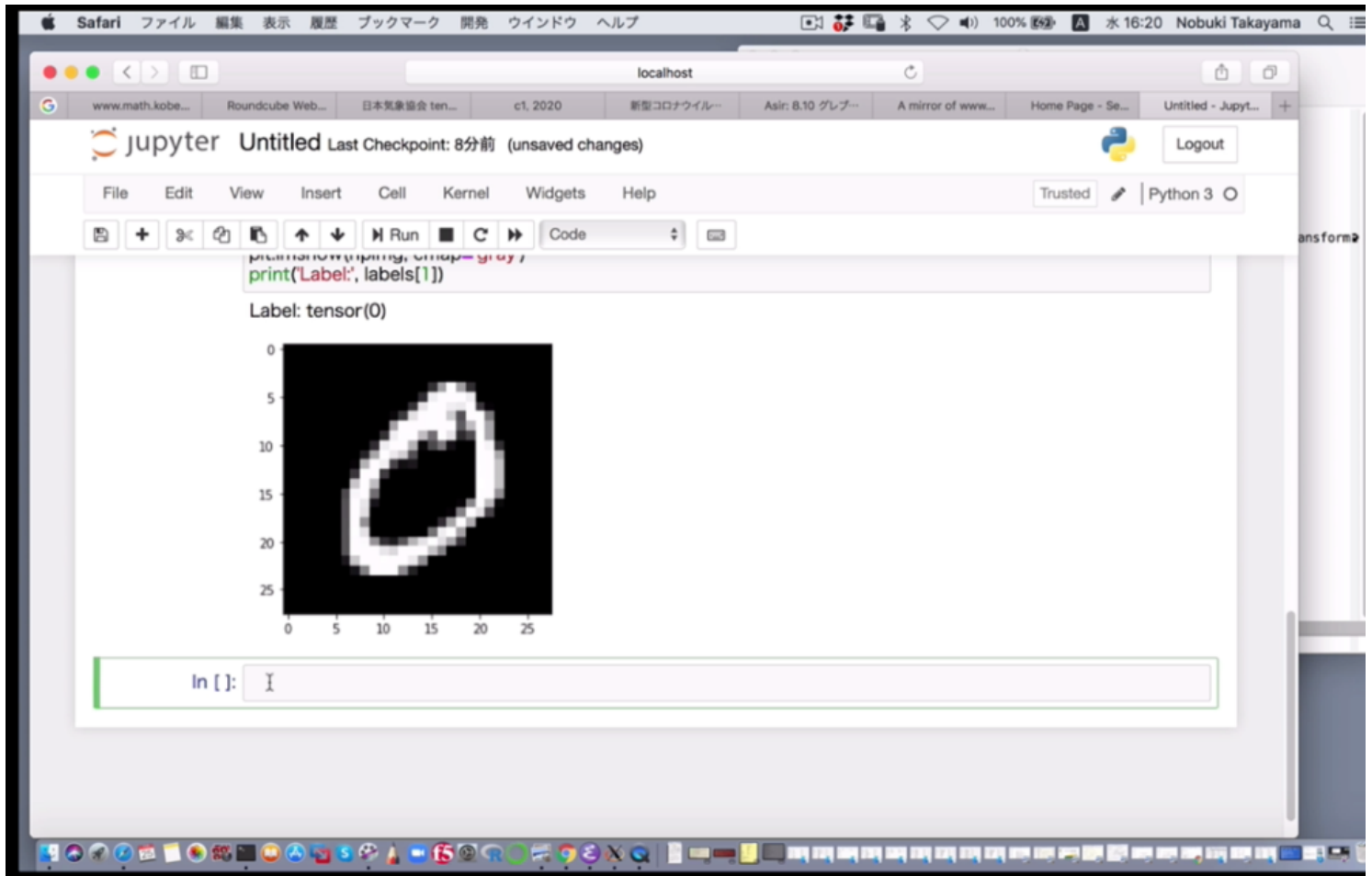
0300.png



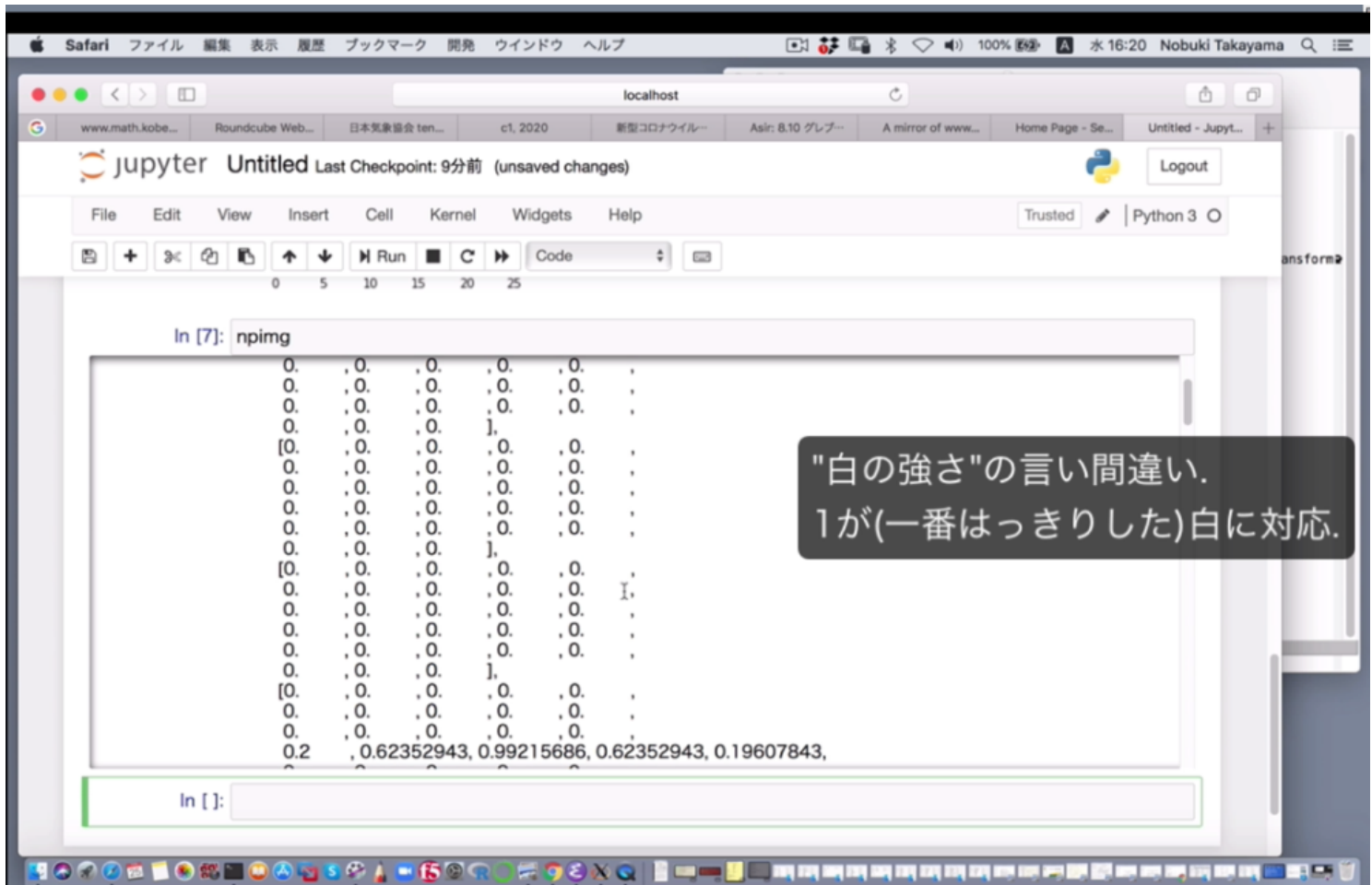
0400.png



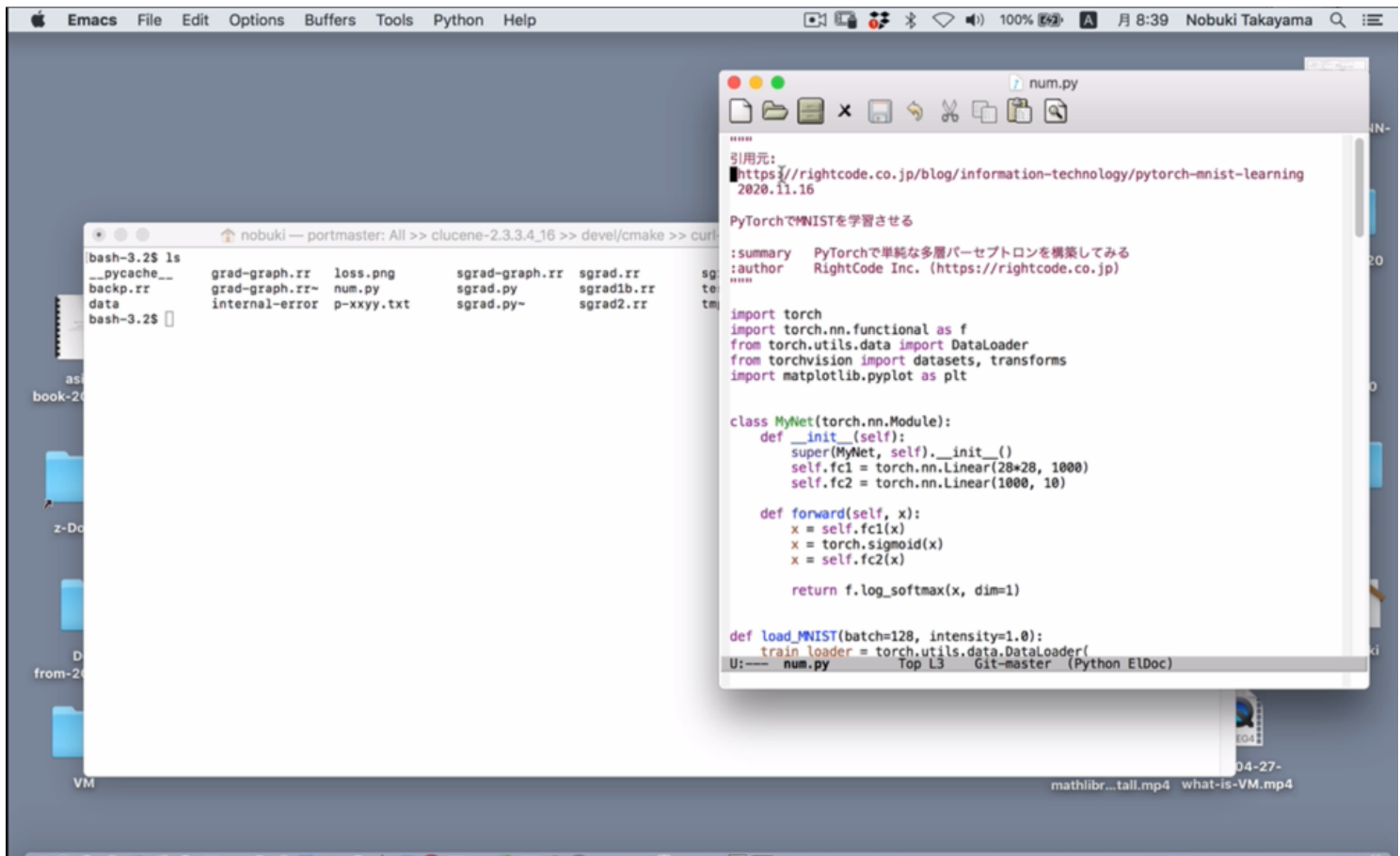
0420.png



0440.png



0600.png



0700.png

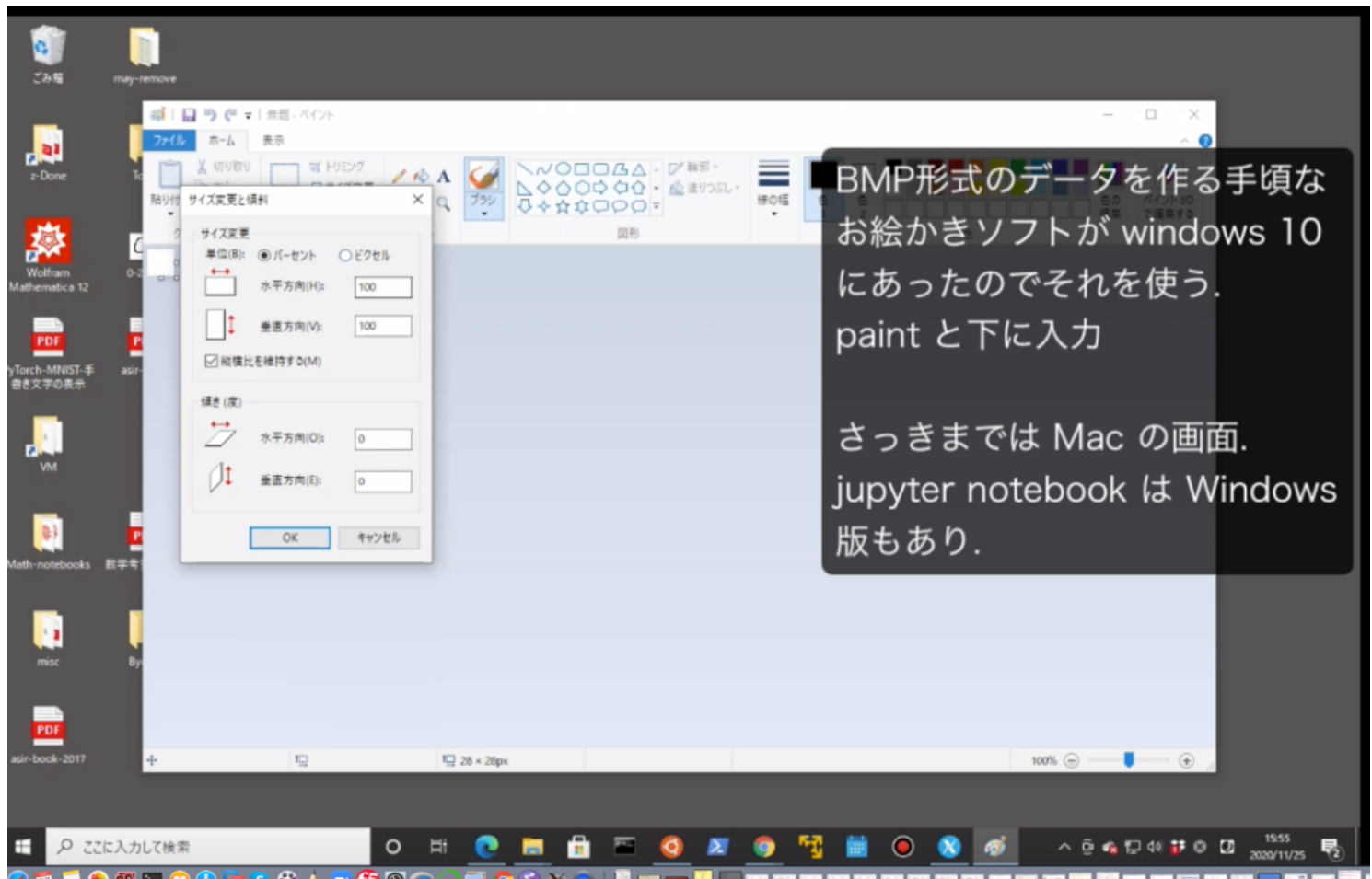
```

nobuki — portmaster: All >> clucene-2.3.3.4_16 >> devel/cmake >> curl-7.62.0 (68/68) — bash - Emacs-x86_64-10_10 — 147x34

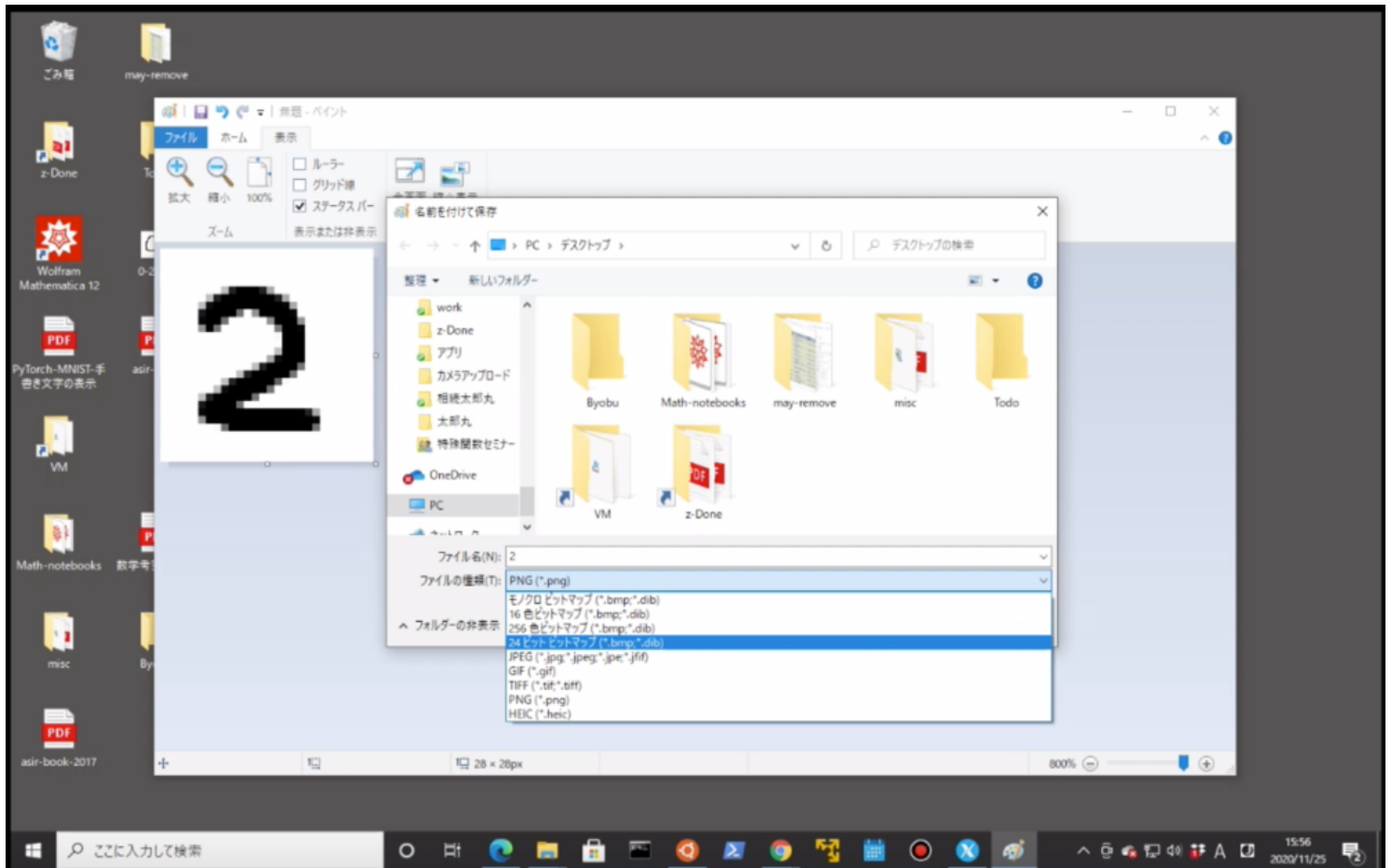
Training log: 5 epoch (23168 / 60000 train. data). Loss: 0.12763309478759766
Training log: 5 epoch (24448 / 60000 train. data). Loss: 0.15563273429870605
Training log: 5 epoch (25728 / 60000 train. data). Loss: 0.10196368945701599
Training log: 5 epoch (27008 / 60000 train. data). Loss: 0.08058089762926102
Training log: 5 epoch (28288 / 60000 train. data). Loss: 0.23345977067947388
Training log: 5 epoch (29568 / 60000 train. data). Loss: 0.08413902670145035
Training log: 5 epoch (30848 / 60000 train. data). Loss: 0.1604592651128769
Training log: 5 epoch (32128 / 60000 train. data). Loss: 0.09811604022979736
Training log: 5 epoch (33408 / 60000 train. data). Loss: 0.06695902347564697
Training log: 5 epoch (34688 / 60000 train. data). Loss: 0.048206791281700134
Training log: 5 epoch (35968 / 60000 train. data). Loss: 0.14217908680438995
Training log: 5 epoch (37248 / 60000 train. data). Loss: 0.10791951417922974
Training log: 5 epoch (38528 / 60000 train. data). Loss: 0.10028531402349472
Training log: 5 epoch (39808 / 60000 train. data). Loss: 0.10346167534589767
Training log: 5 epoch (41088 / 60000 train. data). Loss: 0.1294524371623993
Training log: 5 epoch (42368 / 60000 train. data). Loss: 0.21903829276561737
Training log: 5 epoch (43648 / 60000 train. data). Loss: 0.1023673564195633
Training log: 5 epoch (44928 / 60000 train. data). Loss: 0.09460841864347458
Training log: 5 epoch (46208 / 60000 train. data). Loss: 0.1264665275812149
Training log: 5 epoch (47488 / 60000 train. data). Loss: 0.143461883068008472
Training log: 5 epoch (48768 / 60000 train. data). Loss: 0.05904814600944519
Training log: 5 epoch (50048 / 60000 train. data). Loss: 0.13350121676921844
Training log: 5 epoch (51328 / 60000 train. data). Loss: 0.20717136561870575
Training log: 5 epoch (52608 / 60000 train. data). Loss: 0.14273078739643097
Training log: 5 epoch (53888 / 60000 train. data). Loss: 0.09436320513406062
Training log: 5 epoch (55168 / 60000 train. data). Loss: 0.1541309356689453
Training log: 5 epoch (56448 / 60000 train. data). Loss: 0.0717785581946373
Training log: 5 epoch (57728 / 60000 train. data). Loss: 0.09399757534265518
Training log: 5 epoch (59008 / 60000 train. data). Loss: 0.08480970561504364
Test loss (avg): 0.11448242619633675, Accuracy: 0.9652
{'train_loss': [tensor(0.3576, grad_fn=<NllLossBackward>), tensor(0.1676, grad_fn=<NllLossBackward>), tensor(0.2344, grad_fn=<NllLossBackward>), tensor(0.0652, grad_fn=<NllLossBackward>), tensor(0.0818, grad_fn=<NllLossBackward>)], 'test_loss': [0.2532114329755306, 0.19920841307640075, 0.15569422208070754, 0.13270574288368225, 0.11448242619633675], 'test_acc': [0.9266, 0.94, 0.9528, 0.9592, 0.9652]}
bash-3.2$

```

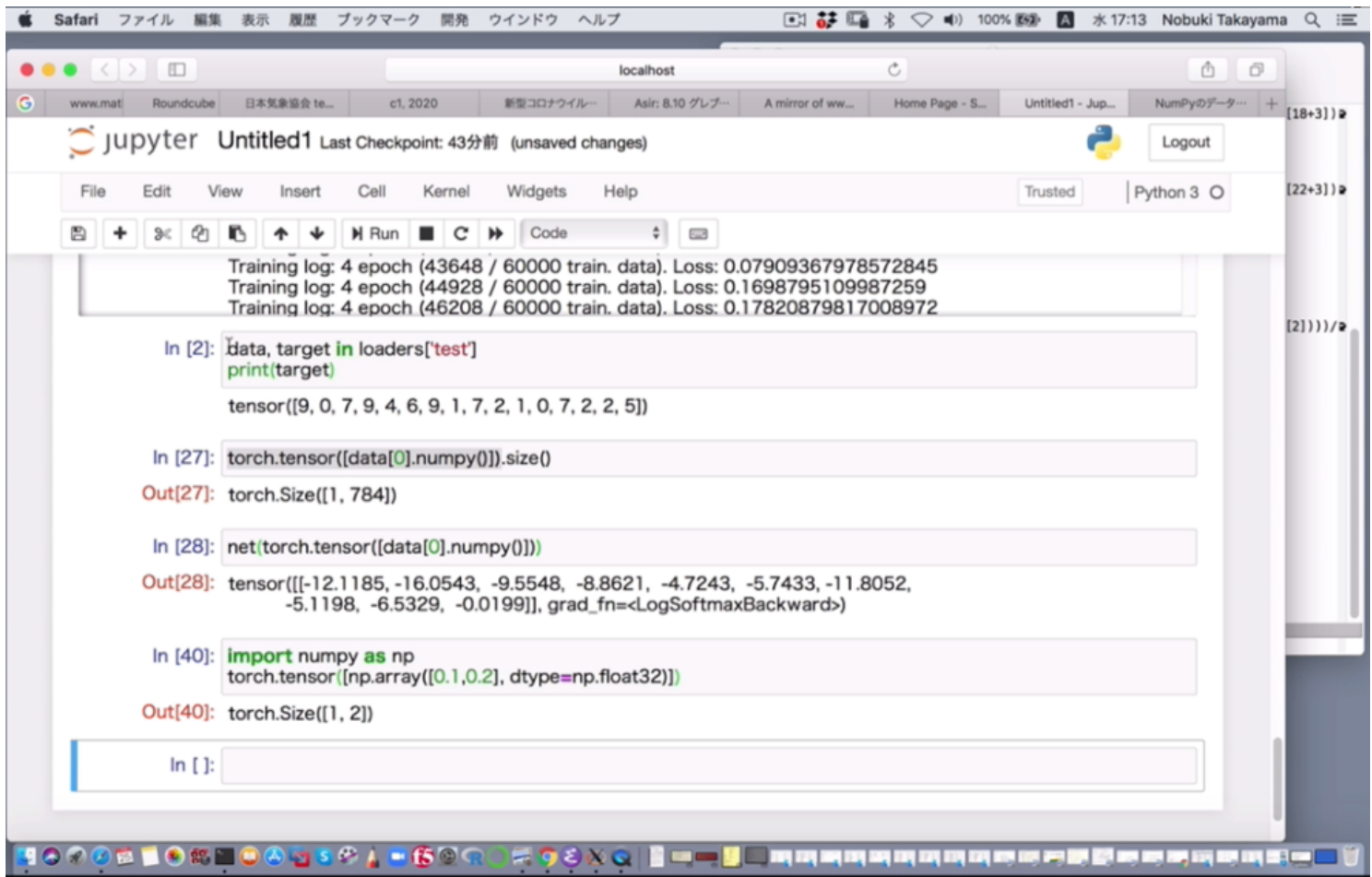
0800.png



0820.png

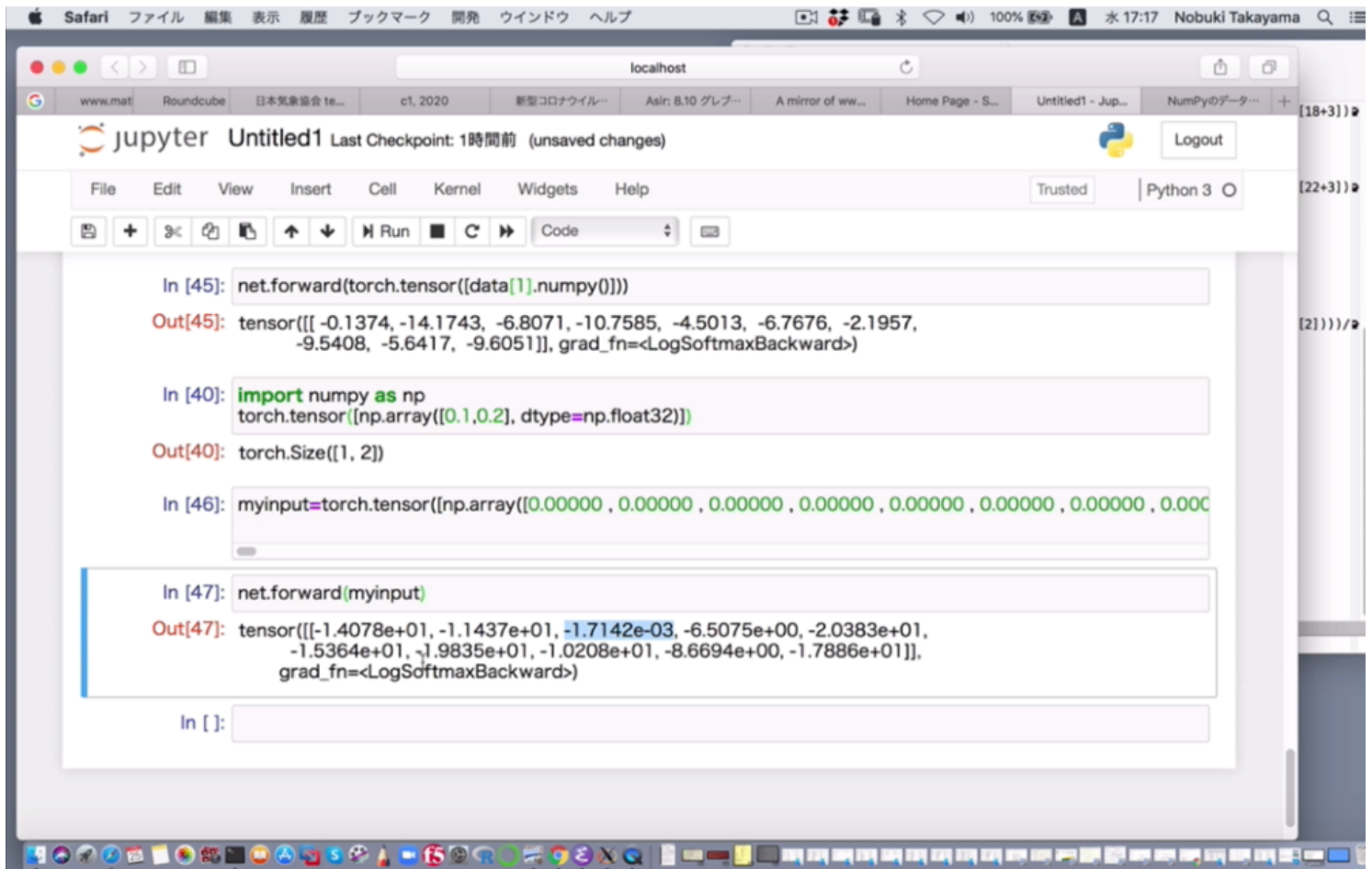


0830.png



0840.png





以下は補足

0860.png

[Movie](#) in the mp4 format.

1000.png

F をきめたい.

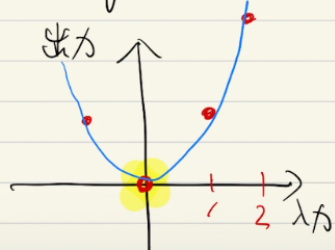
最小値を求める問題に帰着.

λ	F	出力
0		0
1		2
-1		1.9
2		8

$$F(w; x) = wx^2 \quad \leftarrow \text{モデル選択}$$

$$\begin{aligned} \leftarrow \text{loss ft} \\ L(w) &= |F(w, 0) - 0|^2 + |F(w, 1) - 2|^2 \\ &\quad + |F(w, -1) - 1.9|^2 + |F(w, 2) - 8|^2 \end{aligned}$$

loss ft を最小とすような w を求める。



$$w \doteq 2. \Rightarrow F(2; 3) = 18 \text{ 予想}$$

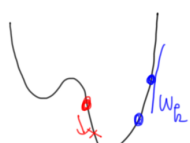
Movie in the mp4 format.

動画: 勾配降下法 (gradient descent) の基礎 (youtube, 限定公開).

10000.png

最適化アルゴリズムの基礎
(Optimization)

$f(w)$ を与え、
の極小値を求めたい。



$f'(w) = 0$, w_0 を与え、 $f'(w_0) > 0$ とする。
 $f'(w_0) < 0$ とする。
 $f'(w_0) = 0$ かつ $f''(w_0) > 0$ とする。
 w_k を答えとする。

$$w_{k+1} = w_k - \epsilon$$

$$w_{k+1} = w_k + \epsilon$$

$$w_{k+1} = w_k - \epsilon f'(w_k)$$

例

$$f(w) = w^3 - w$$

$$= w(w-1)(w+1)$$

$$w_0 = 1 \quad f'(w) = 3w^2 - 1$$

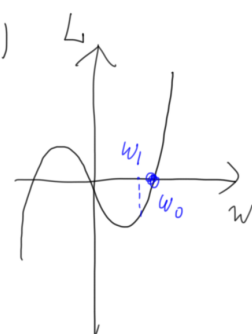
$$f'(w_0) = 3 - 1 = 2$$

$$w_1 = 1 - 2 \cdot 0.1 = 0.8$$

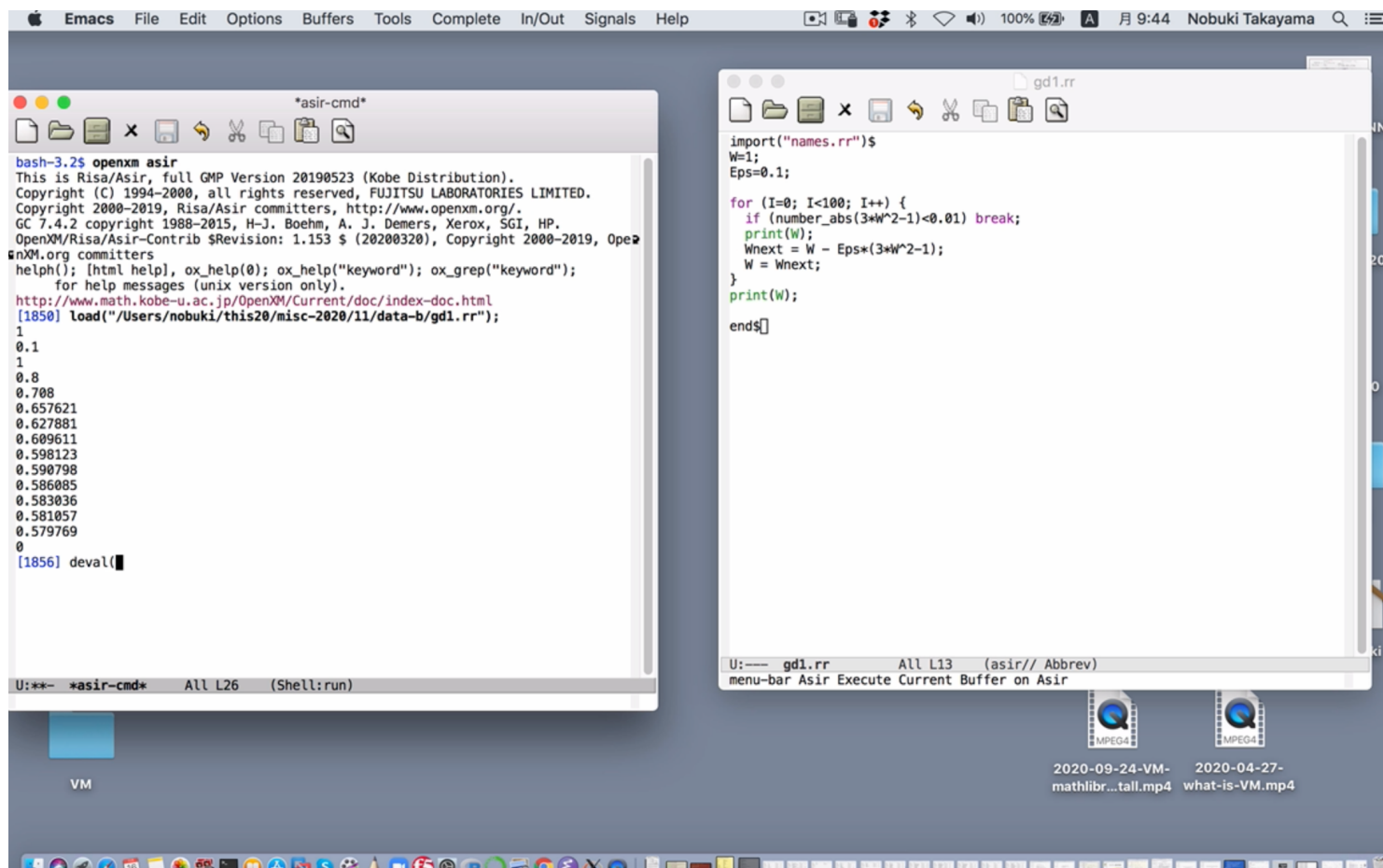
ϵ が小さい Taylor 展開の公式

$$f(w+h) = f(w) + f'(w)h + O(h^2)$$

$$h = -\epsilon f'(w) \text{ とする。}$$



10050.png



10100.png

$$f(w - \varepsilon f'(w))$$

$$= f(w) - \varepsilon (f'(w))^2 + O(\varepsilon^2)$$

(f'(w) ≠ 0)

$f(w_k) \neq 0$

$$f(w_{k+1}) < f(w_k) \text{ のは } \varepsilon \text{ が } \varepsilon + \delta \text{ 以下}$$

2変数で「同」にこだわる。 h_1, h_2 が小さい

$$f(x+h_1, y+h_2)$$

$$= f(x, y) + h_1 \frac{\partial f}{\partial x}(x, y) + h_2 \frac{\partial f}{\partial y}(x, y)$$

$$+ O(h_1^2 + h_2^2)$$

$$x \in w^{(1)}, y \in w^{(2)} \text{ と } x, y$$

$$f(w^{(1)} + h_1, w^{(2)} + h_2)$$

$$= f(w^{(1)}, w^{(2)}) + h_1 \frac{\partial f}{\partial x} + h_2 \frac{\partial f}{\partial y} + O(|h|^2)$$

$$w_k = (w_k^{(1)}, w_k^{(2)})$$

$$f(w_k) \text{ 対 } f(w_k^{(1)}, w_k^{(2)}) \text{ と } \frac{1}{2} \varepsilon^2 \varepsilon \text{ 以下}$$

$$\textcircled{1} \quad w_{k+1} = w_k - \varepsilon \left(\frac{\partial f}{\partial x}(w_k), \frac{\partial f}{\partial y}(w_k) \right)$$

$$f(w_{k+1}) = f(w_k) - \varepsilon \left(\frac{\partial f}{\partial x} \right)^2 - \varepsilon \left(\frac{\partial f}{\partial y} \right)^2 + O(\varepsilon^2)$$

$$\begin{cases} \varepsilon = 10^{-2} & \varepsilon^2 = 10^{-4} \\ \varepsilon = 10^{-3} & \varepsilon^2 = 10^{-6} \end{cases} \quad \begin{cases} h_1 = \varepsilon \frac{\partial f}{\partial x}(w_k) \\ h_2 = \varepsilon \frac{\partial f}{\partial y}(w_k) \end{cases}$$

$$f(w_{k+1}) < f(w_k)$$

10200.png

①を用いて、 $f(w)$ の極小を
求める方法を、

gradient descent とする。

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \text{ } f \text{ の gradient}$$

効率悪い。 $\Rightarrow$

共役 ~~配~~ 法。 $f(x,y)$ が 2 次式なら、
真の解に収束。

conjugate gradient method

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

$f(x,y)$ が 2 次式なら H の固有値はすべて正、異なる。
 $P_k \in \mathbb{R}^2$ $P_i^T H P_j = 0 \quad (i \neq j)$ (仮定)
 $\nwarrow$ H の $w_1, \dots, w_n$ の値

$$w_k \in \mathbb{R}^2$$

$$w_{k+1} = w_k + \alpha_k P_k \quad k=1, 2, \dots$$

$$\alpha_k = - \frac{r_k^T P_k}{P_k^T H P_k}$$

$$r_k = (\nabla \phi)(w_k) \quad \nabla \phi = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

一般の場合

w_2 を新 w_1 とし、上を繰り返す。

α_k は P_k 方向に動かして f が極小となる値に。

H の計算を不要にする方法 いろいろあり。

10300.png

[Movie](#) in the mp4 format.

0100.png

NNを写像 (map) の言葉でいおうと?

集合Aより集合Bへの写像とは、

の元

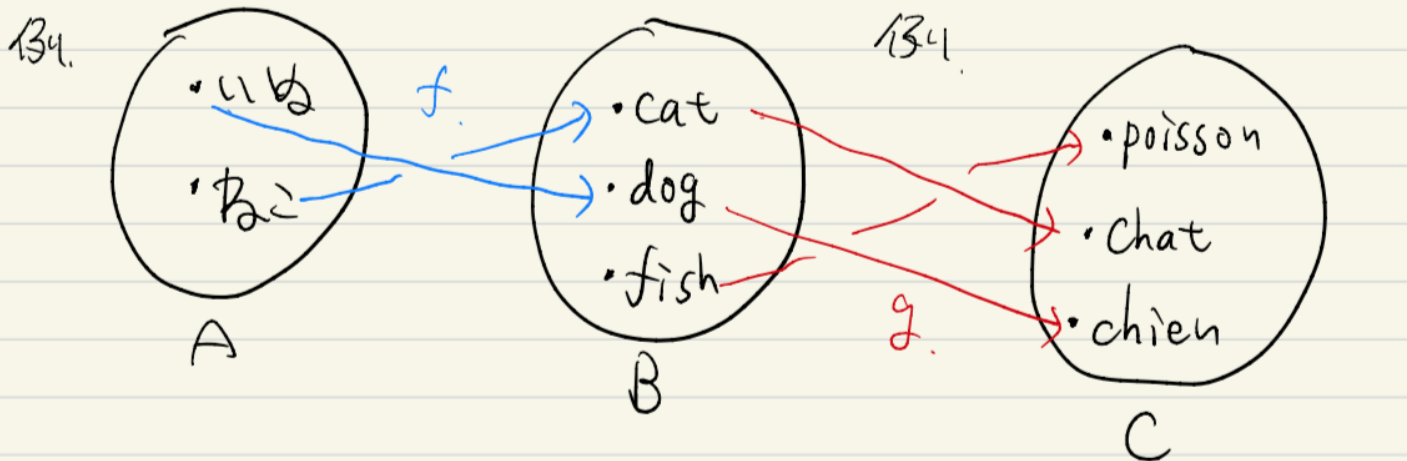
f

任意のAの元に対して、Bをたいてい対応させる規則のこと。

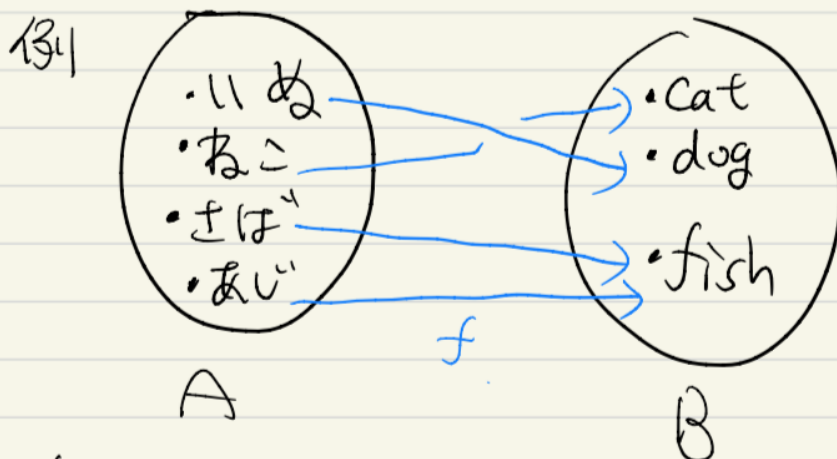
$$a \in A, b \in B$$

$$a \mapsto b$$

$$b = f(a)$$



$$(g \circ f)(\text{いぬ}) = \text{chien}$$



例. $A = B = \text{実数の集合}$. $A \ni x \mapsto x^2 \in B$

0200.png

例. $A=B=$ 平面上の点の集合.

$$A \ni (x, y) \xrightarrow{f} (x+y, x-y) \in B$$

写像の合成

$$f: A \rightarrow B$$

$$g: B \rightarrow C \quad \text{対して}$$

$$A \ni a \xrightarrow{\quad} g(f(a)) \in C$$

$\uparrow$ f と g の合成写像. $g \circ f$ と書く.

例. はし

ハイパーテキスト. とは?

$$\mathbb{R}^n = \{ (x_1, \dots, x_n) \mid x_i \in \mathbb{R} \}$$

たとえば,

$$f: \mathbb{R}^{1000} \rightarrow \mathbb{R}^{10}$$

線型写像

$$f_2: \mathbb{R}^{1000} \xrightarrow{\text{activation function}} \mathbb{R}^{1000}$$

$$f_3: \mathbb{R}^{784} \xrightarrow{\text{線形写像}} \mathbb{R}^{1000}$$

$$f_1 \circ f_2 \circ f_3 \leftarrow \text{10-セグメント}$$

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

$$\frac{d(f \circ g)}{dx} = \frac{df}{dx}(g(x)) \frac{dg}{dx}$$

$$x=p$$

$$q=g(p)$$

$$(f \circ g)'(p) = f'(q) g'(p)$$

$$\frac{d}{dx}(f_1 \circ f_2 \circ f_3)(p) = ?$$

$$q = f_3(p)$$

$$r = f_2(q)$$

$$s = f_1(r)$$

$$\frac{d}{dx}(f_1 \circ (f_2 \circ f_3))(p)$$

$$\begin{aligned} & \downarrow \\ & = f'_1(v) \cdot (f_2 \circ f_3)'(p) \\ & = f'_1(v) \cdot f'_2(q) \cdot f'_3(p) \\ & \text{backpropagation.} \end{aligned}$$

微分積分で習う方法はダメ?

そんなことはない. 数学ソフトウェアのアルゴリズムには微分積分で習う方法をさらに精緻にしたものがある. 解ける問題のサイズは小さいが正確な答えをだせる.

1. 一変数代数方程式, Schönhage algorithm. pari/gp の roots.

2. 多変数, Gröbner basis, たとえば

<http://www.math.kobe-u.ac.jp/Asir/index-ja.html>

3. Numerical homotopy method, たとえば

<http://homepages.math.uic.edu/~jan/download.html>

[1982] F;

$5x^{10}-5x^8+x-5$

[1983] G=diff(F,x);

$50x^9-40x^7+1$

[1984] pari(roots,G);

[-0.91501003876985439820584647933955481421

0.66087367610411118725987469773911708804

0.86231598027599305196534276125095206872

虚部のある解は省略.]