Building Deep Learning applications with Keras:
Machine Learning Review

Felipe Salvatore
https://felipessalvatore.github.io/

Lucas Moura
http://lmoura.me/

June 15, 2018

Machine Learning Basics

Machine learning recap

A machine learning algorithm is an algorithm capable of using data to solve a task

Regression

$x_1, \ldots, x_N$

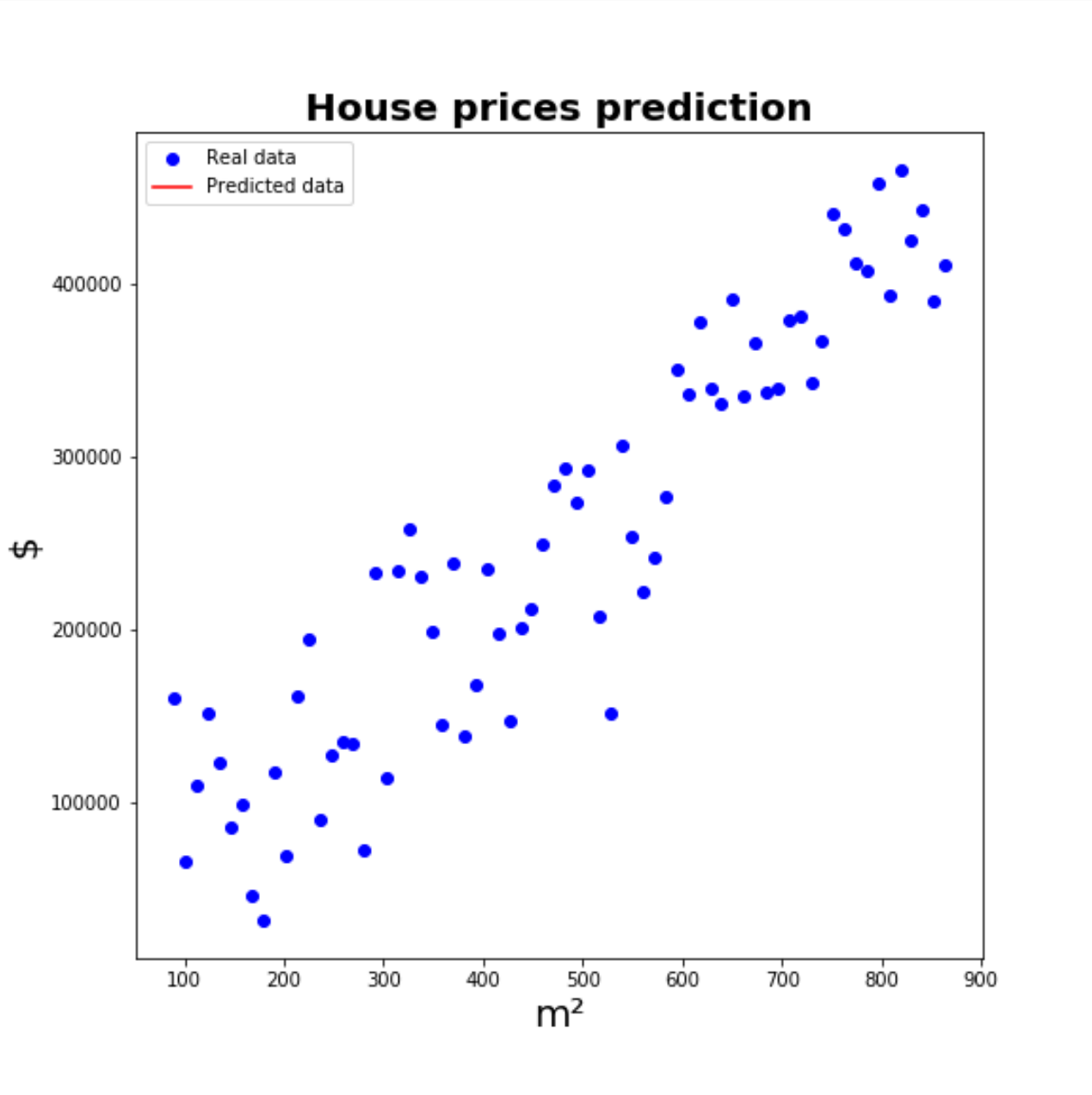
$y_1, \ldots, y_N$

Classification

$x_1, \ldots, x_N$

$L_1, \ldots, L_N$

Regression

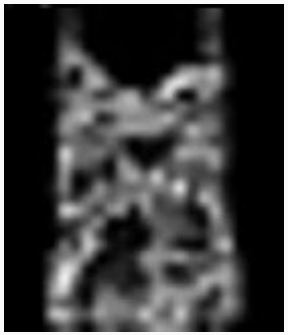

House prices prediction
Real data
Predicted data
400000
300000
200000
100000
$
100
200
300
400
500
600
700
800
900
m²

Regression


House prices prediction
Real data
Predicted data
600000
500000
400000
300000
200000
100000
$
200
400
600
800
1000
1200
m²

Classification: Fashion MNIST [6]

$y =$ **shirt**



$y =$ **dress**

$y =$ **handbag**

$y =$ **sneaker**

Machine learning recipe

- A learning task.
- A family of models.
- A cost function.
- An optimization algorithm.

Machine learning recipe

- A learning task. Fashion MNIST
- A family of models. Logistic regression
- A cost function. Cross entropy
- An optimization algorithm. Stochastic Gradient Descent

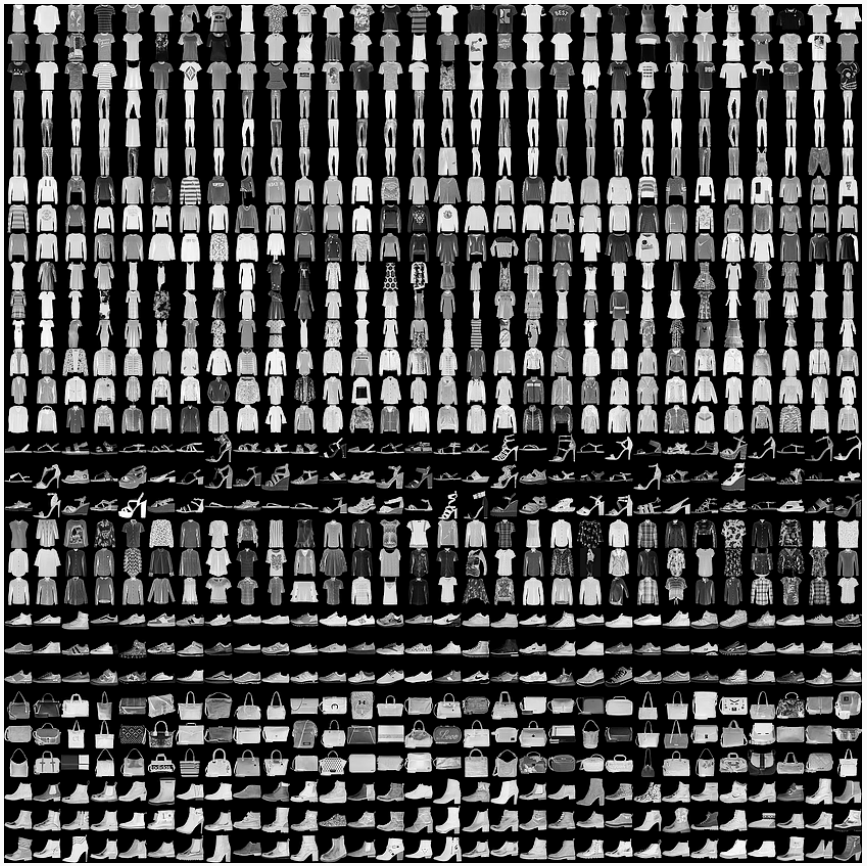
Splitting the data

- train data
- valid data
- test data

Splitting the data

- train data used to learn model's parameters
- valid data used to change hyperparameters
- test data model evaluation

Fashion MNIST



- Training set: 60k
- Test set: 10k
- Images: 28×28 grayscale.
- 10 classes: (shirt, dress, sandals, ...)

Fashion MNIST

28×28

$$\boldsymbol{x} = \begin{bmatrix} 0.34 \\ \vdots \\ 0.06 \end{bmatrix}$$

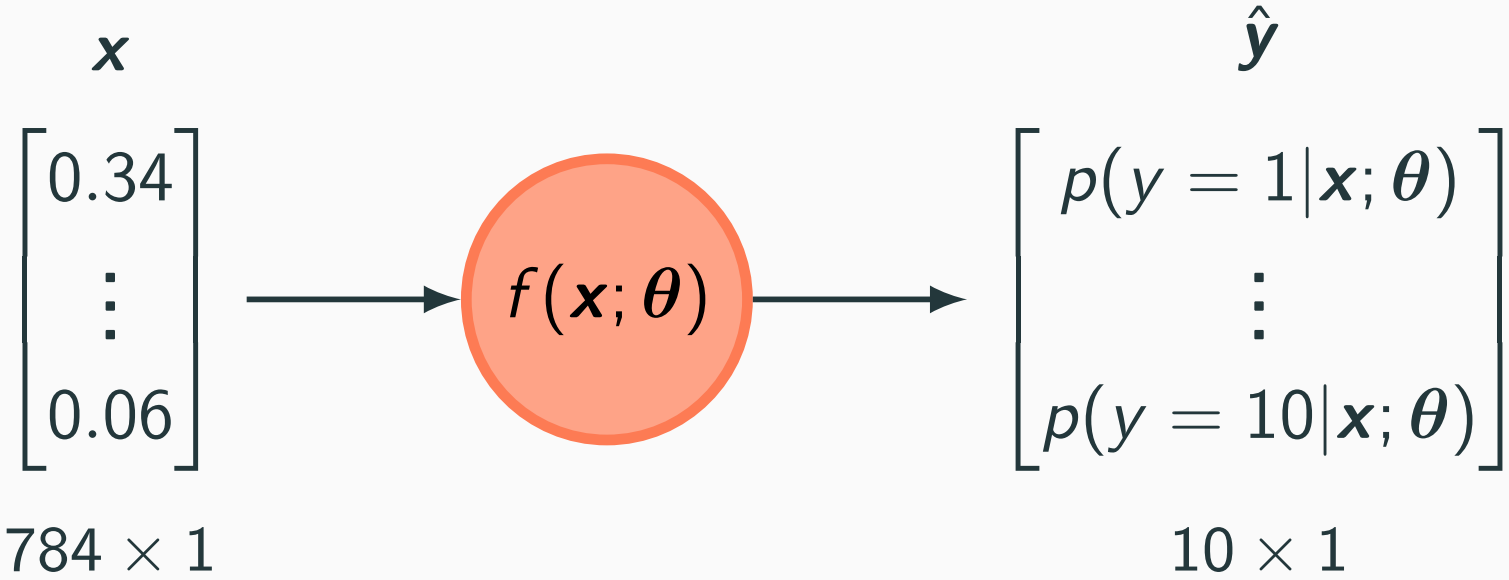
784×1

$y = 1$

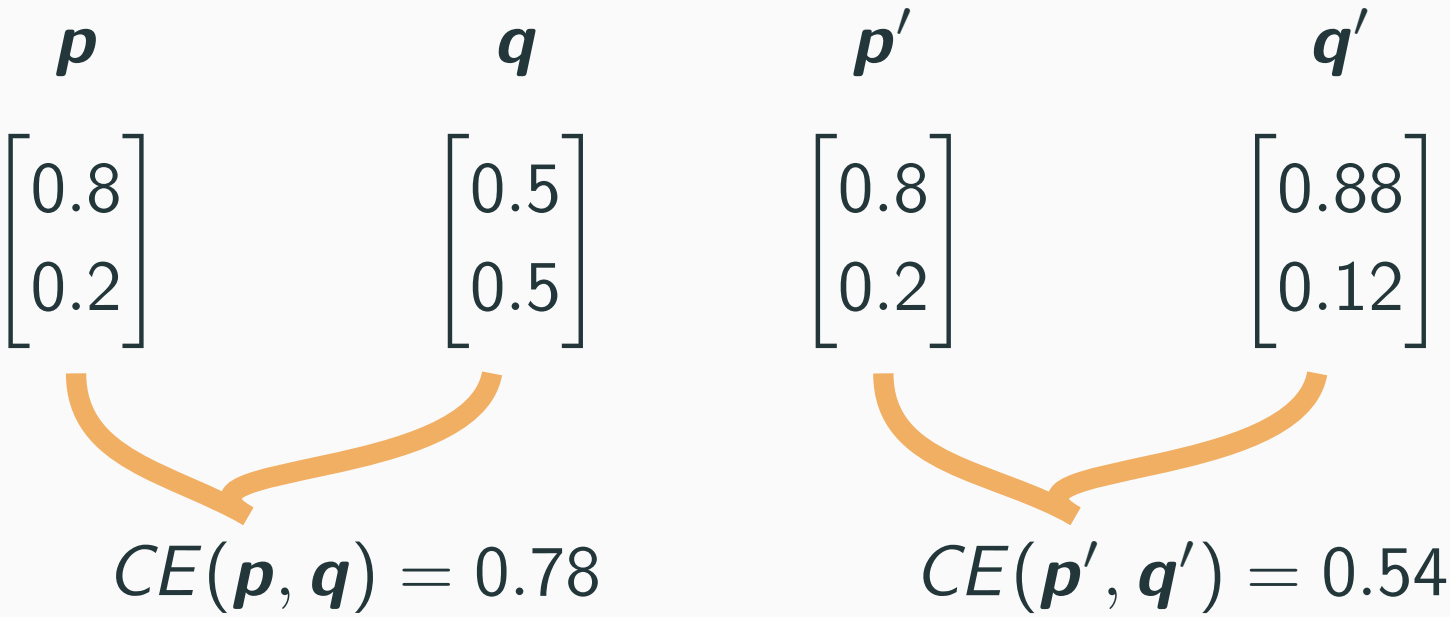
$$\boldsymbol{y} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$$

10×1

Multilabel classification



Cross entropy



$$CE(\boldsymbol{p}, \boldsymbol{q}) = -\sum_i \boldsymbol{p}_i \log \boldsymbol{q}_i$$

Stochastic gradient descent

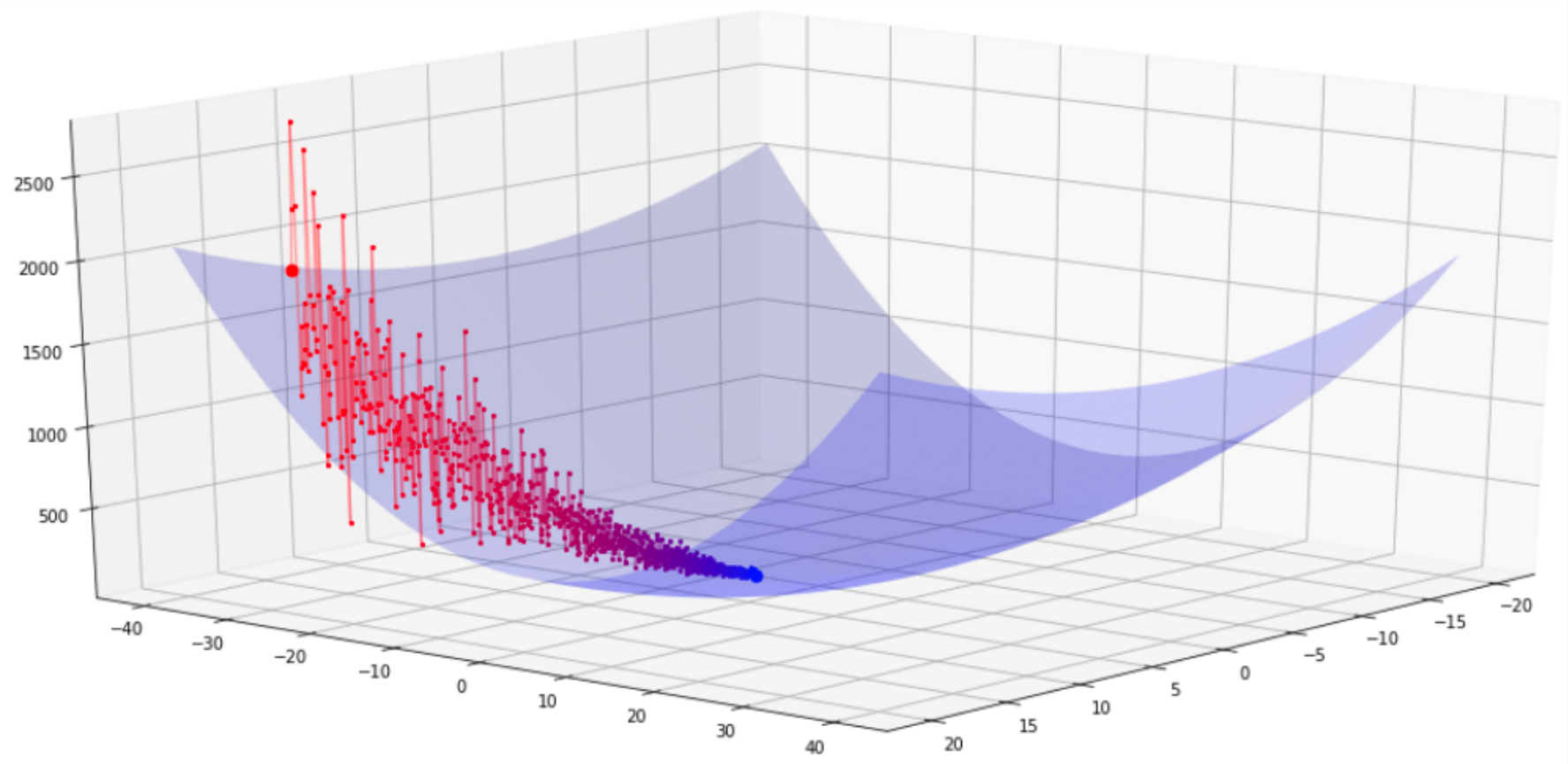
- Sample a minibatch of m examples ($m \ll N$) from the training data $(\boldsymbol{x}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{x}^{(m)}, y^{(m)})$
- Compute the gradient estimate: $\nabla_{\boldsymbol{\theta}} \hat{J}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \frac{1}{m} \sum_{i=1}^{m} L(\boldsymbol{y}^{(i)}, \hat{\boldsymbol{y}}^{(i)})$
- Apply update : $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old} - \alpha \nabla_{\boldsymbol{\theta}} \hat{J}(\boldsymbol{\theta})$

Stochastic gradient descent

- Sample a minibatch of m examples ($m \ll N$) from the training data $(\boldsymbol{x}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{x}^{(m)}, y^{(m)})$
- Compute the gradient estimate: $\nabla_{\boldsymbol{\theta}} \hat{J}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \frac{1}{m} \sum_{i=1}^{m} L(\boldsymbol{y}^{(i)}, \hat{\boldsymbol{y}}^{(i)})$
- Apply update : $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old} - \alpha \nabla_{\boldsymbol{\theta}} \hat{J}(\boldsymbol{\theta})$

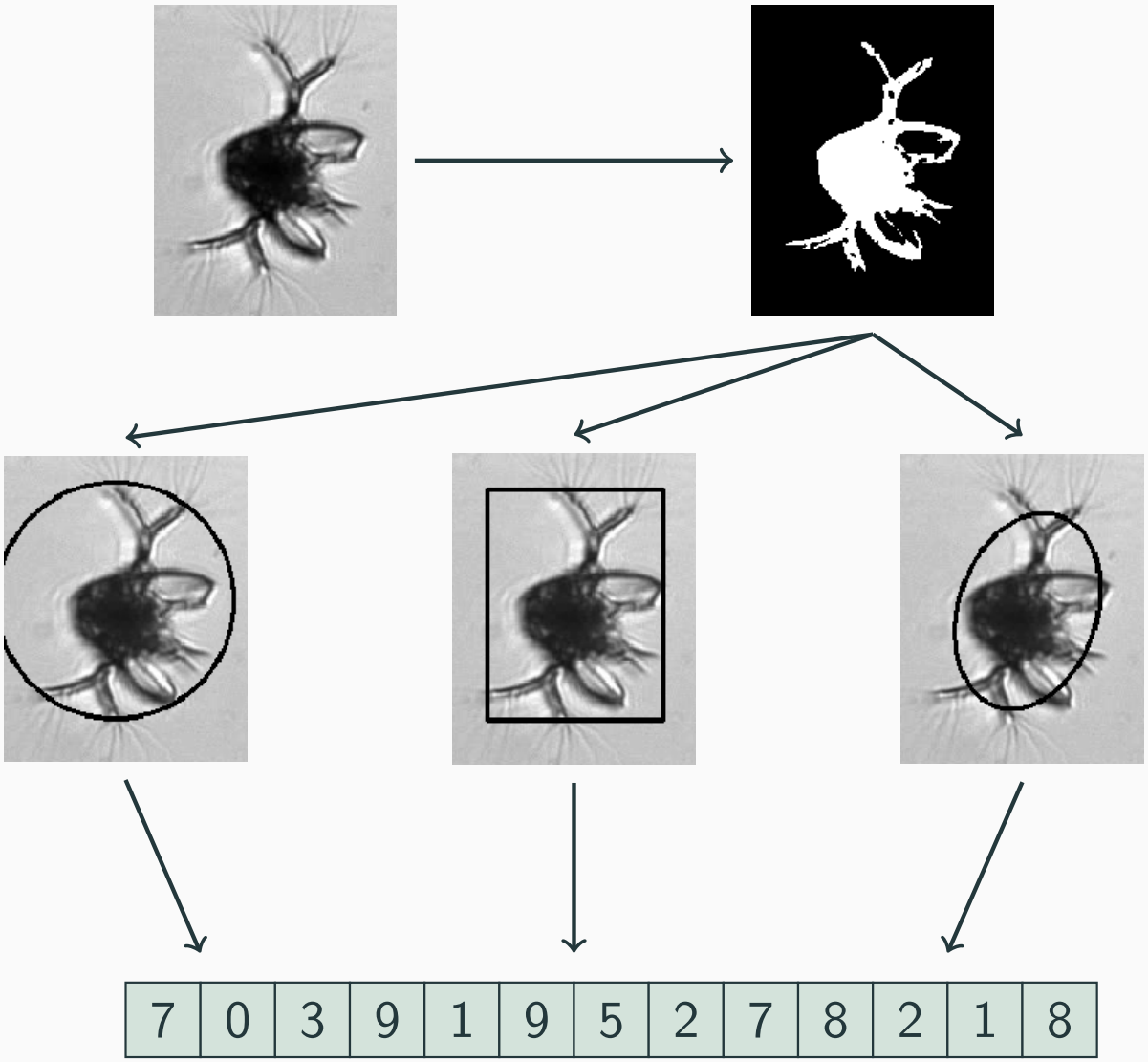
We use the term epoch to describe when an entire dataset is passed through the machine learning model once.

Stochastic gradient descent



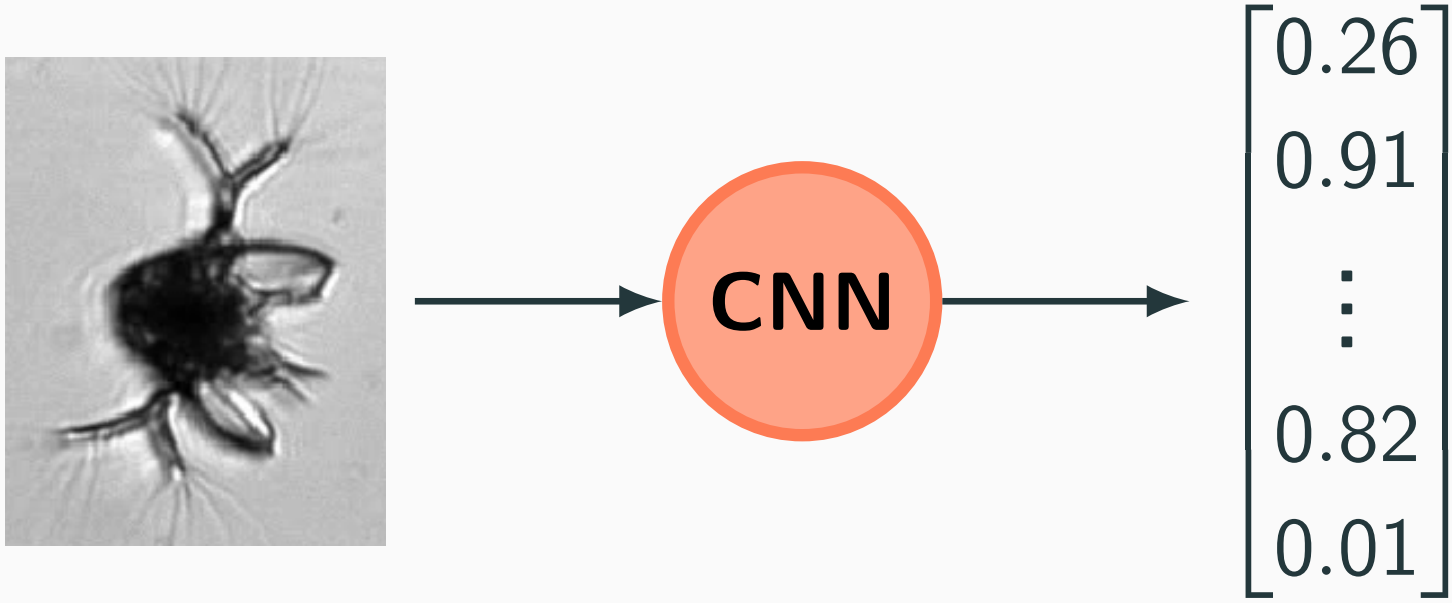
Deep Learning

Feature Engineering: Image Classification



vector describing different image statistics

Representation Learning



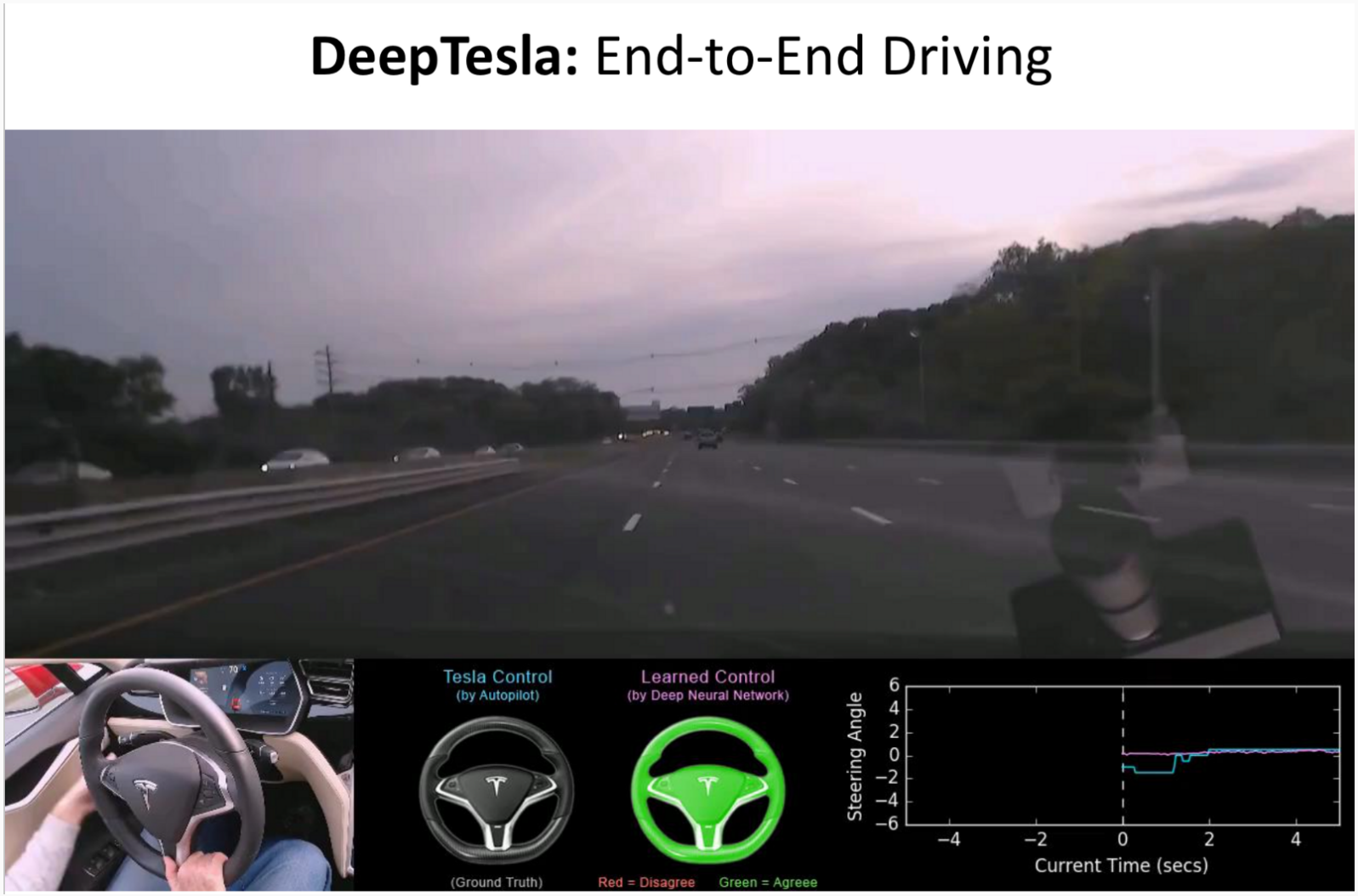
Deep Learning allows us to delegate the feature engineering task to the optimization algorithm

Deep Learning

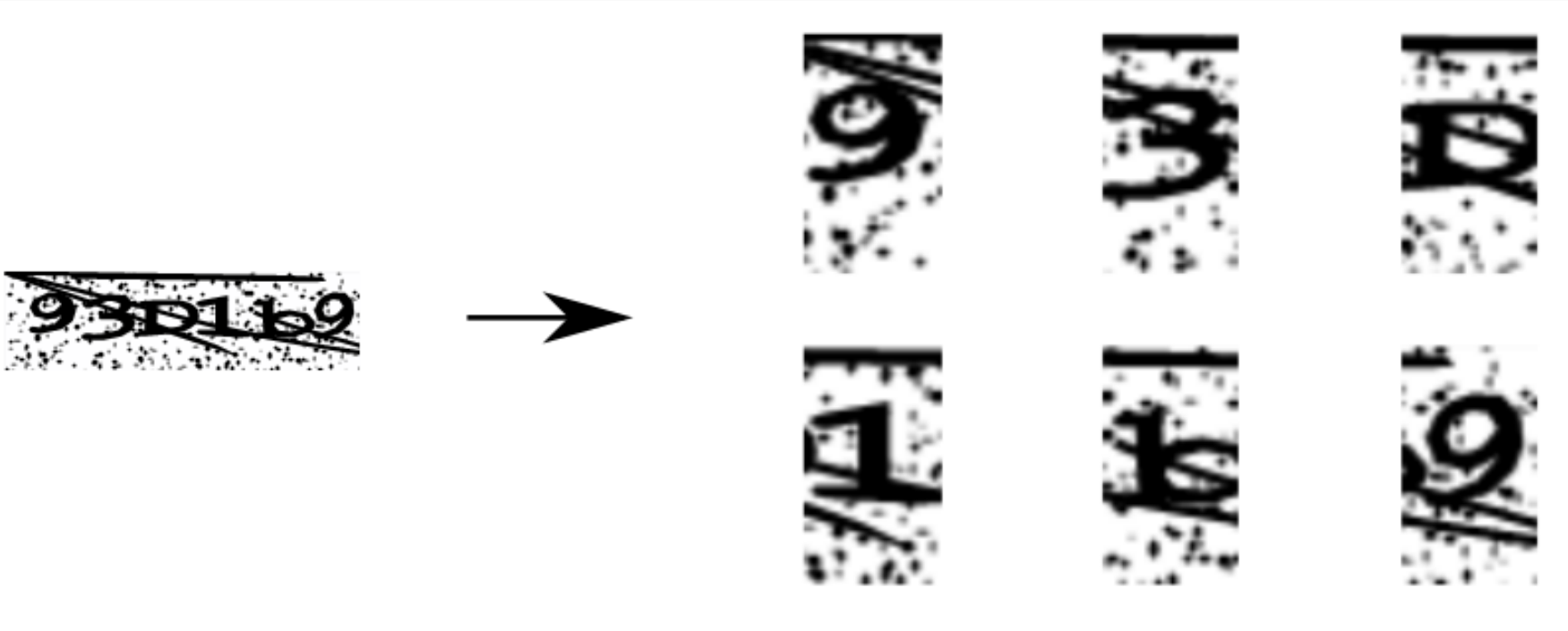
"The solution is to allow computers to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts. (...) This approach avoids the need for human operators to formally specify all the knowledge that the computer needs. (...) If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach deep learning."

I. Goodfellow, Y. Bengio, A. Courville – **Deep Learning** (2017)

Applications: Self-Driving Car [2]



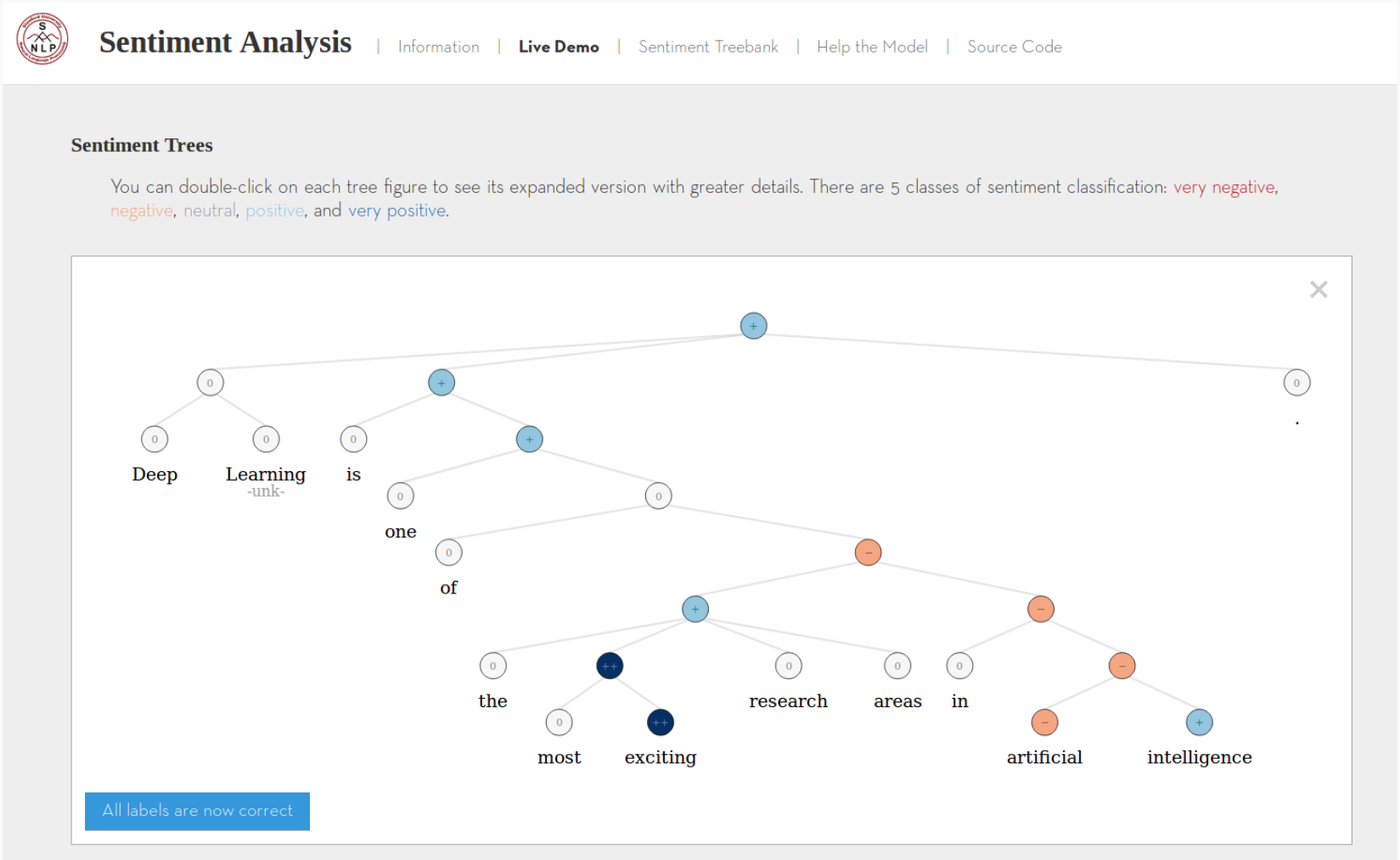
Applications: Breaking Captchas [1]



Applications: Text Generation [3]

Só celebridade
Só moleque chique
Os mais top da cidade
De cidade da cidade
E o menino que cresceu e geral fala o que elas
quer o tempo nós quer adrenalina pra nós é verão
e também pode vim que eu quero ver o baile
e o guime é do bom e as melhores mulheres
Eu vou dormir lá em baixo na casa dos macho
Pode ter certeza que lá eu me acho
Eu sei que eu me acho
Eu vou dormir lá em cima
Na casa das prima
Lá o uísque é do bom

Applications: Sentiment analysis [4]



References I

- Deep learning drops: breaking captcha.
  `https://towardsdatascience.com/deep-learning-drops-breaking-captcha-20c8fc96e6a3`.
- Deep tesla.
  `https://selfdrivingcars.mit.edu/deeptesla/`.
- Mc neural: O funkeiro artificial.
  `https://medium.com/@lucasmoura_35920/mc-neural-o-funkeiro-artificial-ab6fbedc9771`.
- Stanford sentiment treebank.
  `http://nlp.stanford.edu:8080/sentiment/rntnDemo.html`.
- I. Goodfellow, Y. Bengio, and A. Courville.
  Deep Learning.
  MIT Press, 2017.

References II


H. Xiao, K. Rasul, and R. Vollgraf.
Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.