Supervised Learning

CSC 461: Machine Learning

Fall 2020

Prof. Marco Alvarez
University of Rhode Island

Supervised Learning Setup

Components of (supervised) learning

- Input space $\mathcal{X}$
- Output space $\mathcal{Y}$
- Data instance $x \in \mathcal{X}, y \in \mathcal{Y}$
  - ✓ is a pair (x,y)
- Data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$
  - ✓ is a set of data instances
- Hypothesis $g : \mathcal{X} \mapsto \mathcal{Y}, g \in \mathcal{H}$

Example

Problem:
automatically tagging email messages as spam (1) or ham (0)

Input Space?
assume every email is represented as a fixed-length vector of 10 features

Output Space

Data

- Samples are assumed to be **independent and identically distributed** from the same probability distribution (**i.i.d**)

$$\mathscr{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

in general $\mathscr{X} = \mathbb{R}^d$

$$(x_i, y_i) \sim P$$

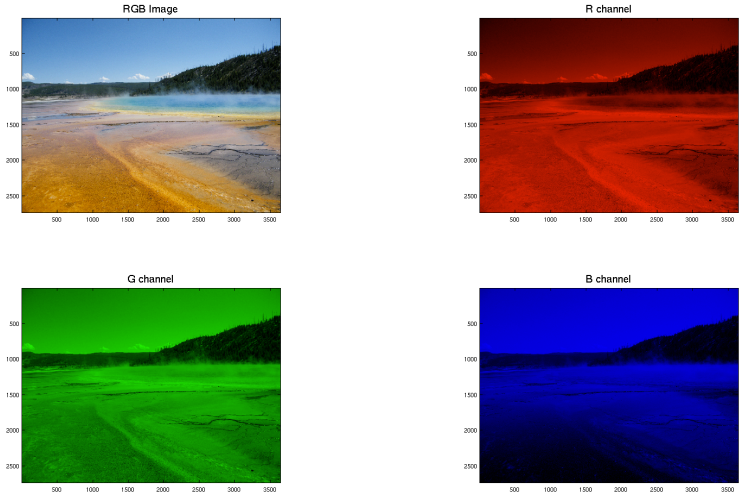
unknown

What are Feature Vectors?

$$(x_i, y_i) \sim P$$

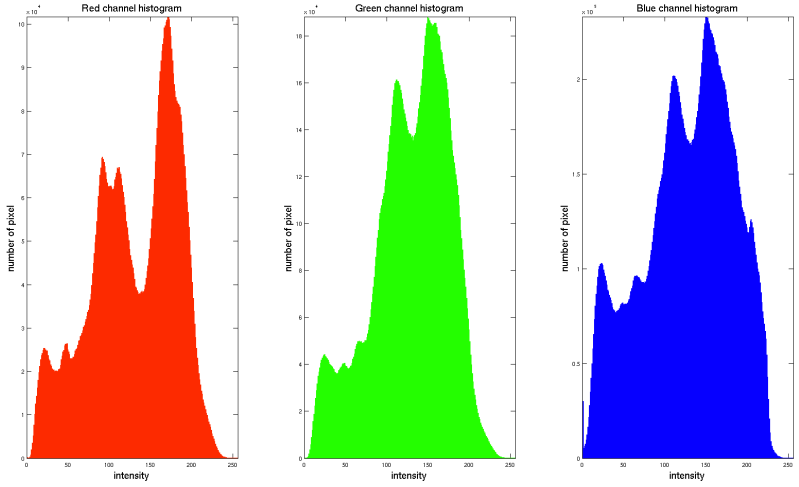
unknown

Feature Vectors



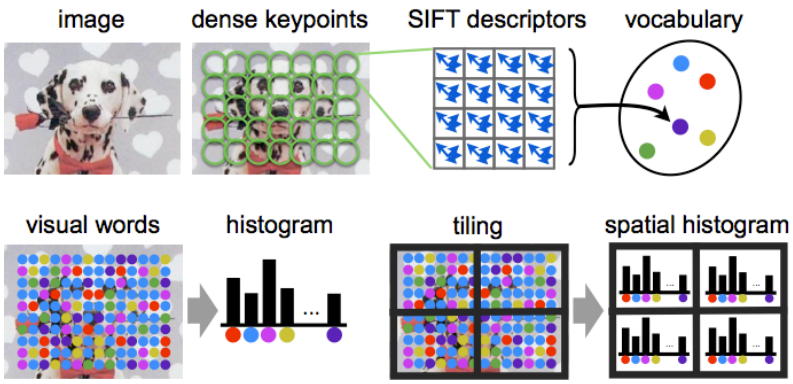
http://www.sci.utah.edu/~acoste/uou/Image/project1/Arthur_COSTE_Project_1_report.html

Feature Vectors



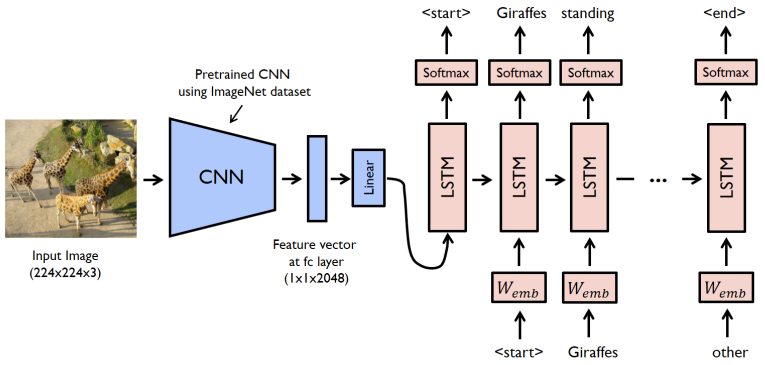
http://www.sci.utah.edu/~acoste/uou/Image/project1/Arthur_COSTE_Project_1_report.html

Feature Vectors



https://www.di.ens.fr/willow/teaching/recvis13/assignment2/

Feature Vectors

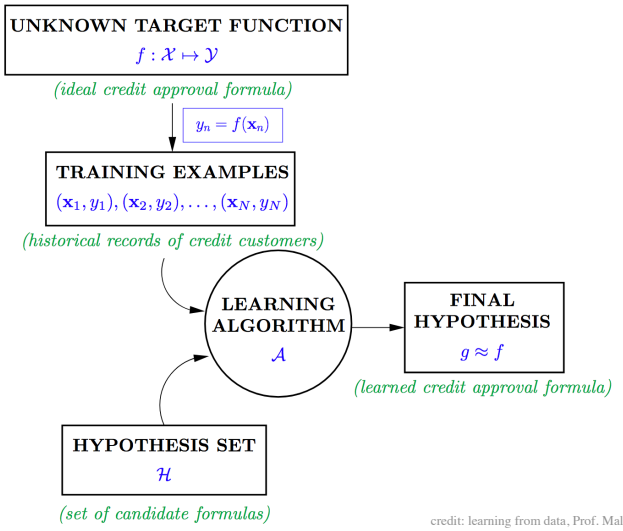


https://www.analyticsvidhya.com/blog/2018/04/solving-an-image-captioning-task-using-deep-learning/

Supervised learning

Binary classification	$\mathcal{Y} = \{0,1\}$ $\mathcal{Y} = \{-1,+1\}$
Multiclass classification	$\mathcal{Y} = \{0,1,\ldots,k-1\}$
Regression	$\mathcal{Y} = \mathbb{R}$
Structure prediction	structured objects

Learning setup



credit: learning from data, Prof. Malik Magdon-Ismail

Example

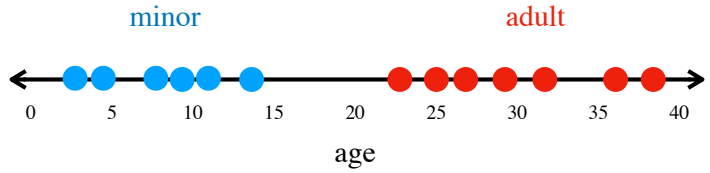
$h_1 \in \mathscr{H}$

$h_2 \in \mathscr{H}$

$\cdots$

can you define the hypothesis space?

how to pick a hypothesis that makes you happy?



Loss Functions

- 0/1 Loss $\mathscr{L}_{0/1}(h, \mathscr{D}) = \frac{1}{n} \sum_{(x_i, y_i) \in \mathscr{D}} I(h(x_i) \neq y_i)$

  indicator function

- Squared Loss $\mathscr{L}_{sq}(h, \mathscr{D}) = \frac{1}{n} \sum_{(x_i, y_i) \in \mathscr{D}} (h(x_i) - y_i)^2$

- Absolute Loss $\mathscr{L}_{abs}(h, \mathscr{D}) = \frac{1}{n} \sum_{(x_i, y_i) \in \mathscr{D}} |h(x_i) - y_i|$

Consider this classifier …

$$g(x) = \begin{cases} y_i & \text{if } x = x_i \text{ for } (x_i, y_i) \in \mathscr{D} \\ y_1 & \text{otherwise} \end{cases}$$

what is the 0/1 loss for training data? square loss? absolute loss?

are you happy with this classifier? why?

What is the goal of (supervised) learning?

- A function (**classifier/regressor**) that best approximates target function

For $g \in \mathscr{H}$ and $\forall (x_i, y_i) \sim P$, we want $g(x) \approx f(x)$

search and **optimization** (to **minimize expected loss**)

Role of Training and Testing

Example on Iris Dataset (next lecture)

From the UCI ML repository …

Iris Data Set

Download: Data Folder, Data Set Description

Abstract: Famous database; from Fisher, 1936

Data Set Characteristics:	Multivariate	**Number of Instances:**	150	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	4	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	2096826

Attribute Information:

- sepal length in cm
- sepal width in cm
- petal length in cm
- petal width in cm

Class Information:

- Iris Setosa
- Iris Versicolour
- Iris Virginica

Example using Iris dataset

https://colab.research.google.com/drive/1pDkSHtAG1A2kCKrPY1CNDtVdKj1f_wK3