Algorithmic Operation Research

Homework 3

Nikolaos Galanis - sdi1700019
Pantelis Papageorgiou - sdi1700115
Maria-Despoina Siampou - sdi1600151

November 3, 2019

What is Image Classification?

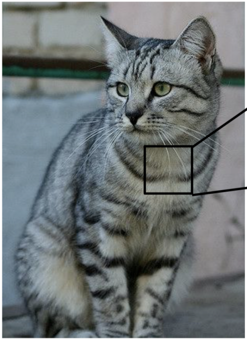
Defintion

Image classification refers to the task of extracting information classes from a multiband raster image. The resulting raster from image classification can be used to create thematic maps.

(assume given set of discrete labels)
{dog, cat, truck, plane, ...}

$\longrightarrow$ cat

The Problem: Semantic Gap




```
[[105 112 108 111 104  99 106  99  96 103 112 119 104  97  93  87]
 [ 91  98 102 106 104  79  98 103  99 105 123 136 110 105  94  85]
 [ 76  85  90 105 128 105  87  96  95  99 115 112 106 103  99  85]
 [ 99  81  81  93 120 131 127 100  95  98 102  99  96  93 101  94]
 [106  91  61  64  69  91  88  85 101 107 109  98  75  84  96  95]
 [114 108  85  55  55  69  64  54  64  87 112 129  98  74  84  91]
 [133 137 147 103  65  81  80  65  52  54  74  84 102  93  85  82]
 [128 137 144 140 109  95  86  70  62  65  63  63  60  73  86 101]
 [125 133 148 137 119 121 117  94  65  79  80  65  54  64  72  98]
 [127 125 131 147 133 127 126 131 111  96  89  75  61  64  72  84]
 [115 114 109 123 150 148 131 118 113 109 100  92  74  65  72  78]
 [ 89  93  90  97 108 147 131 118 113 114 113 109 106  95  77  80]
 [ 63  77  86  81  77  79 102 123 117 115 117 125 125 130 115  87]
 [ 62  65  82  89  78  71  80 101 124 126 119 101 107 114 131 119]
 [ 63  65  75  88  89  71  62  81 120 138 135 105  81  98 110 118]
 [ 87  65  71  87 106  95  69  45  76 130 126 107  92  94 105 112]
 [118  97  82  86 117 123 116  66  41  51  95  93  89  95 102 107]
 [164 146 112  80  82 120 124 104  76  48  45  66  88 101 102 109]
 [157 170 157 120  93  86 114 132 112  97  69  55  70  82  99  94]
 [130 128 134 161 139 100 109 118 121 134 114  87  65  53  69  86]
 [128 112  96 117 150 144 120 115 104 107 102  93  87  81  72  79]
 [123 107  96  86  83 112 153 149 122 109 104  75  80 107 112  99]
 [122 121 102  80  82  86  94 117 145 148 153 102  58  78  92 107]
 [122 164 148 103  71  56  78  83  93 103 119 139 102  61  69  84]]
```

What the computer sees

An image is just a big grid of numbers between [0, 255]:

e.g. 800 x 600 x 3
(3 channels RGB)

Parametric Approach: Linear Classifier

3072x1

$$\underbrace{\boxed{f(x,W)}}_{10\times 1} = \underbrace{\boxed{W}}_{10\times 3072}\overbrace{\boxed{x}}^{3072\times 1} + \boxed{b} \quad 10\times 1$$

10x1 10x3072

Image

Array of **32x32x3** numbers (3072 numbers total)

f(**x**,**W**)

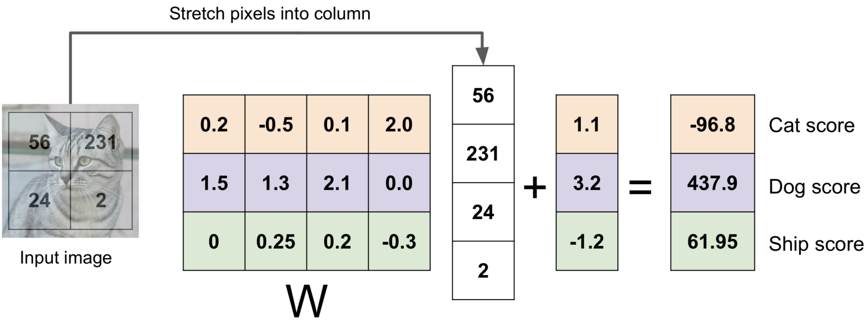
10 numbers giving class scores

W

parameters or weights

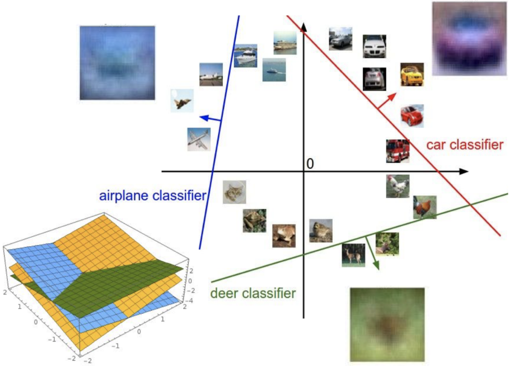
Example

Example with an image with 4 pixels, and 3 classes (cat/dog/ship)



But what is this thing doing?

Interpreting a Linear Classifier



$$f(x, W) = Wx + b$$

Array of **32x32x3** numbers
(3072 numbers total)

Support Vector Machine

Motivation

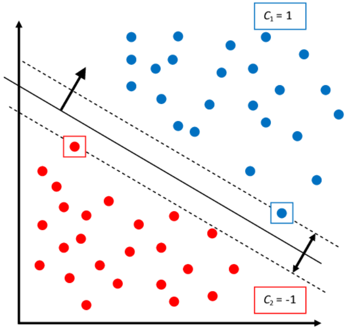
Motivation: Given some data points each belonging to one of two classes - decide which class a new data point will be in.

Linear Classifier

Data points are viewed as a **p**-dimensional vectors. We want to know whether we can separate such points with a **(p-1)**-dimensional hyperplane. This is called a **linear classifier**.

Support Vector Machine

An Introduction



Given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new examples.

Support Vector Machine

Linear SVM

We are given a training dataset of n points of the form: $(\vec{x}_1 y_1)..(\vec{x}_n y_n)$

$$\text{where the } y_i = \begin{cases} 1 & \text{for positive sample} \\ -1 & \text{for negative sample} \end{cases} \tag{1}$$

Any hyperplane can be written as the set of points $\vec{x}$ satisfying:

$$\vec{w} \cdot \vec{x} - b = 0 \tag{2}$$

Support Vector Machine

Hard - Margin

If the training data is linearly separable, we can select two parallel hyperplanes that separate the two classes of data, so that the distance between them is **as large as possible**.

With a normalized or standardized dataset, these hyperplanes can be described by the equations:

$$\begin{cases} \vec{w} \cdot \vec{x} - b = 1 \\ \vec{w} \cdot \vec{x} - b = -1 \end{cases} \tag{3}$$

Support Vector Machine

Hard - Margin

To prevent data points from falling into the margin, we add the following constraint: for each i either:

$$\begin{cases} \vec{w} \cdot \vec{x} - b \geq 1 \\ \vec{w} \cdot \vec{x} - b \leq -1 \end{cases} \iff \vec{y}(\vec{w} \cdot \vec{x} - b) \geq 1 \iff \boxed{\vec{y}(\vec{w} \cdot \vec{x} - b) - 1 \geq 0} \tag{4}$$

These constraints state that each data point must lie on the correct side of the margin.

Support Vector Machine

Hard - Margin

Geometrically, the distance between these two hyperplanes is:

$$(\vec{x}_+ - \vec{x}_-) \cdot \frac{\vec{w}}{\|\vec{w}\|} = \frac{\vec{w}\vec{x}_+ - \vec{w}\vec{x}_-}{\|\vec{w}\|} = \frac{1 - b + 1 + b}{\|\vec{w}\|} = \boxed{\frac{2}{\|\vec{w}\|}} \qquad (5)$$

Support Vector Machine

Hard - Margin

Summing up, we get the above optimization problem:

Minimize $\|\vec{w}\|$ subject to $y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad \forall 1 \leq i \leq n$. The $\vec{w}$ and b that solve this problem determine our classifier.

Pattern Classification via Linear Programming

Application

Our intention

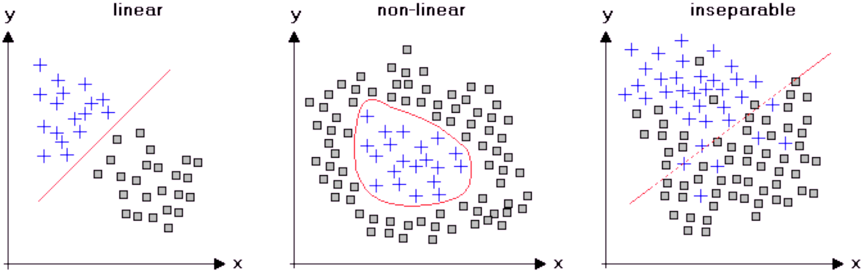
Demonstrate how Linear Programming can be used to perform linear classification

Given 2 sets $H = H^1, \ldots, H^h \subseteq \mathrm{R}^n$ and $M = H^1, \ldots, H^h \subseteq \mathrm{R}^n$, we want to:

1. check if H and M are linearly separable
2. if so, find a separating hyperplane

Linear separability

Definition



Linear separability

Alternative definition

Theorem

Two sets $H \subseteq \mathrm{R}^n$ and $M \subseteq \mathrm{R}^n$ are said to be linearly separable if $\exists a \in \mathrm{R}^n, b \in \mathrm{R} : H \subseteq x \in \mathrm{R}^n : a^T > b$ and $M \subseteq x \in \mathrm{R}^n : a^T \leq b$

Lemma

Two sets $H = H^1, \ldots, H^h \subseteq \mathrm{R}^n$ and $M = H^1, \ldots, H^h \subseteq \mathrm{R}^n$ are linearly seperable if and only if there exists $a \in \mathrm{R}^n$ and $b \in \mathrm{R}$ such that $a^T H^i - b \geq 1$ and $a^T M^j - b \leq -1 \; \forall i \in \{1, \ldots, h\}, \; \forall j \in \{1, \ldots, m\}$

Linear separability

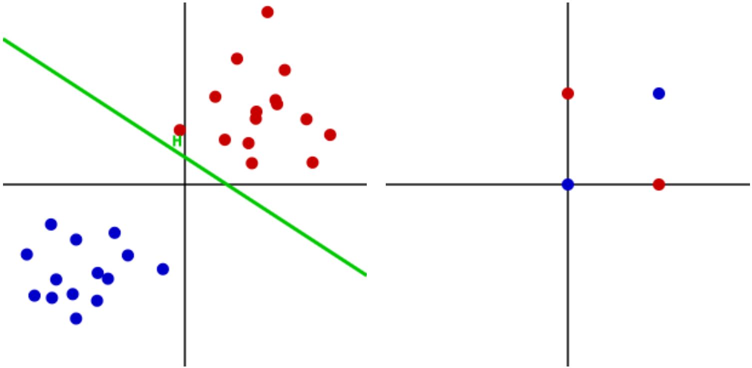
Linear classification in dimension 1

Linear separability

Linear classification in dimension 1

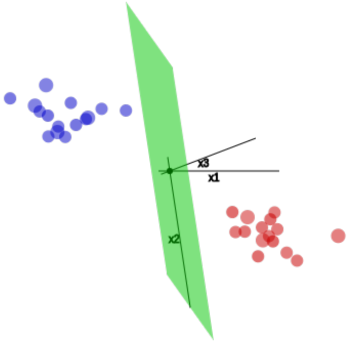
Linear separability

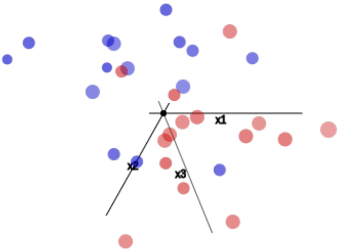
Linear classification in dimension 2



Linear separability

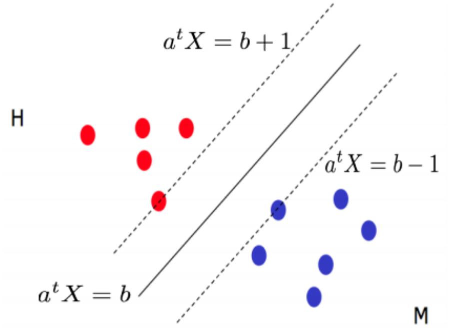
Linear classification in dimension 3





Linear separability

Alternative definition



Pattern Classification via Linear Programming

Appilication

Goal

Show how linear programming can be used to solve problems (1) and (2) as shown above for 2 given sets $H = H^1, \ldots, H^h \subseteq \mathrm{R}^n$ and $M = H^1, \ldots, H^h \subseteq \mathrm{R}^n$

Linear problem

$$\min_{y,z,a,b} \frac{1}{h}[y_1 + y_2 + \cdots + y_h] + \frac{1}{m}[z_1 + z_2 + \cdots + z_m]$$

Linear problem

Constraints for y_i

- $y_i \geq -a^T H^i + b + 1, \ i = 1, \ldots, h$
- $y_i \geq 0, \ i = 1, \ldots, h$

Constraints for z_i

- $z_i \geq -a^T M^i + b + 1, \ j = 1, \ldots, h$
- $z_i \geq 0, \ j = 1, \ldots, h$

where $y \in \mathrm{R}^h, \ z \in \mathrm{R}^m, \ a \in \mathrm{R}^n, \ b \in \mathrm{R}.$
In the following, we will refer to the above program as LP.

Useful theorems to solve LP

Theorem (1)

H and M are linearly separable if and only if the optimal value of LP is 0.

Theorem (2)

If H and M Are linearly separable and y*,z*,a*,b* is an optimal solution of LP, then $f(x) = {a^*}^T x + b^*$ is a separating hyperplane.

Example

Linearly Separable Case

Initial data

- $H = \{(0,0),(1,0)\} \subset \mathrm{R}^2$ and $M = \{(0,2),(1,2)\} \subset \mathrm{R}^2$
- So, $h = 2, \ m = 2$

Driven by the above theorems, we can associate the data with the following:

Linear problem

$$\min_{y,z,a,b} \frac{1}{2}[y_1 + y_2] + \frac{1}{2}[z_1 + z_2]$$

Maths

We can proceed as following:

- $y_1 \geq -[a_1, a_2] \begin{bmatrix} 0 \\ 0 \end{bmatrix} + b + 1 = b + 1$
- $y_2 \geq -[a_1, a_2] \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b + 1 = -a_1 + b + 1$
- $z_1 \geq [a_1, a_2] \begin{bmatrix} 0 \\ 2 \end{bmatrix} + b + 1 = 2a_2 - b + 1$
- $z_1 \geq [a_1, a_2] \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b + 1 = a_1 + 2a_2 - b + 1$

where $y_i \geq 0, \;\; i = 1, 2$ and $z_i \geq 0, \;\; i = 1, 2$

Optimal Solution

Optimal solution for the LP

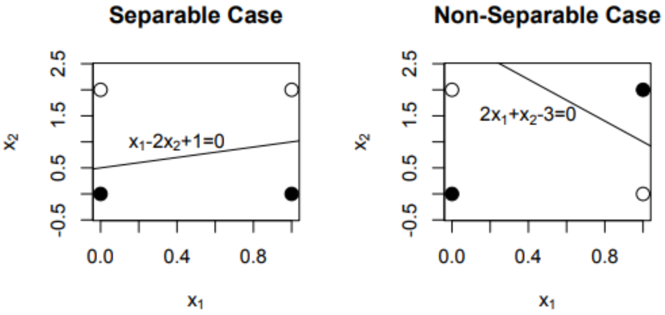
- $y_1 = y_2 = z_1 = z_2 = 0$
- $a^T = [1, -2], \; b = -1$

Separating hyperplane

Considering the theorem (2), we get: $x_1 - 2x_2 + 1 = 0$

But what is this thing doing?

Interpreting a Linear Classifier



References

- Carnegie Mellon University Department of Statistics Private Web Area: https://www.stat.cmu.edu/
- Stanford Department of Computer Science: http://cs231n.stanford.edu/2017/
- MIT School of Engineering: http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf