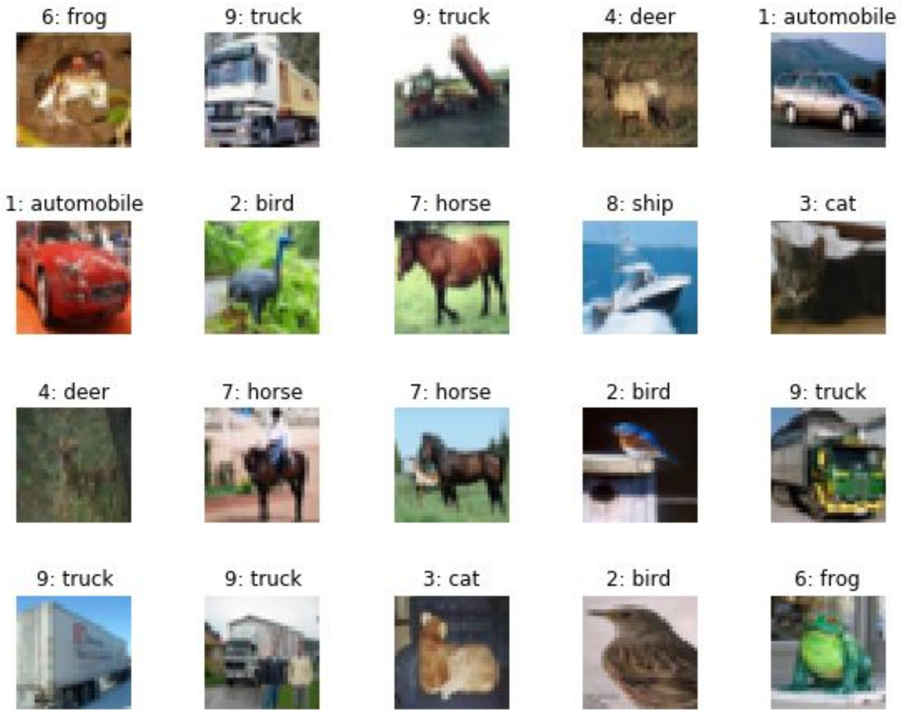
SVM for binary classification
in CIFAR-10 dataset

Matteo Marre’ Brunenghi - a.a. 2018-2019

CIFAR-10

Dimensionality:

- 32*32 px * 3 channels per image
- 5000 training images per class
- 10 classes
- 1000 test images



Dimensionality overview

Frogs vs Planes

32 * 32 * 3 = 3072

RGB to greyscale

32 * 32 = 1024

Histogram of oriented gradients

16 * 16 = 256

Data recentering, scaling
PCA

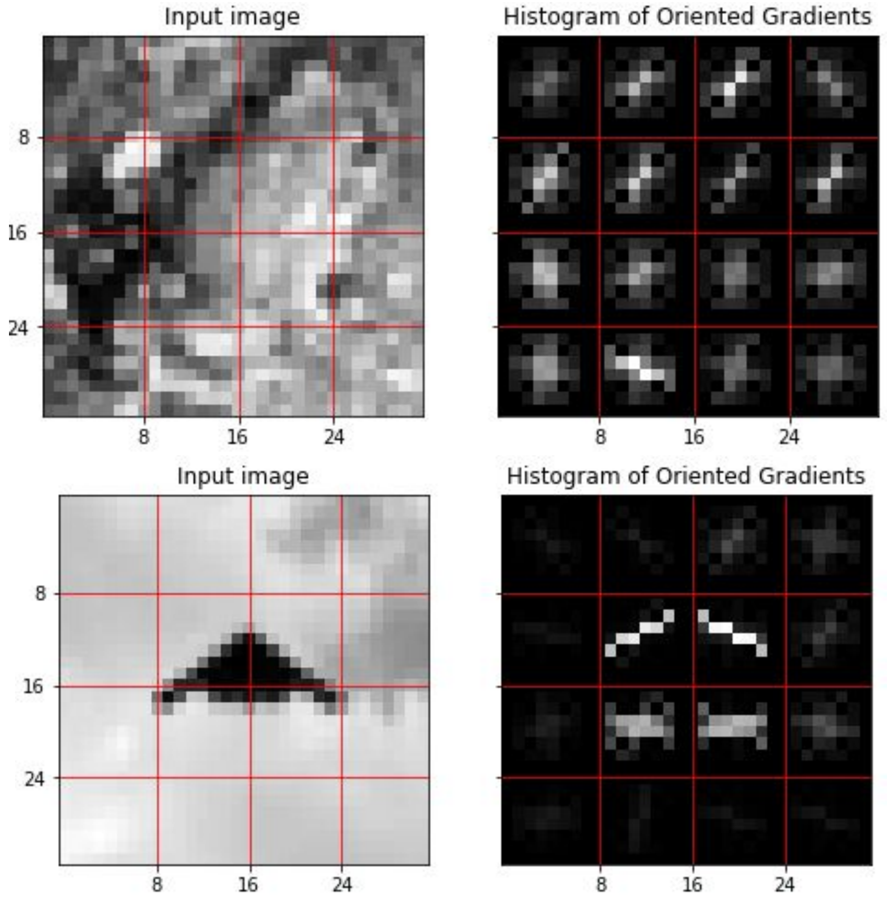
176

SVM

Solution

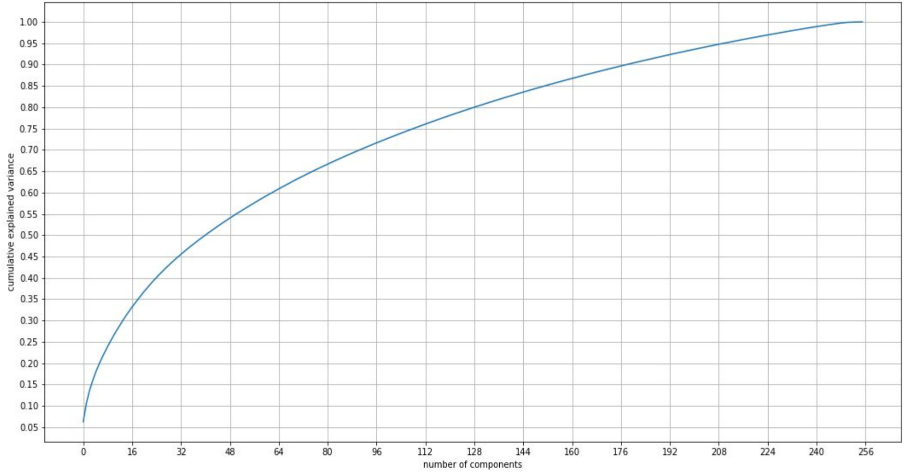
Preprocessing: Histogram of oriented gradients

- computing the gradient image in row and col with Sobel filter
- computing gradient histograms
- flattening into a feature vector



Scaling and PCA

Cumulative explained variance 90%
with 176 features



SVM?

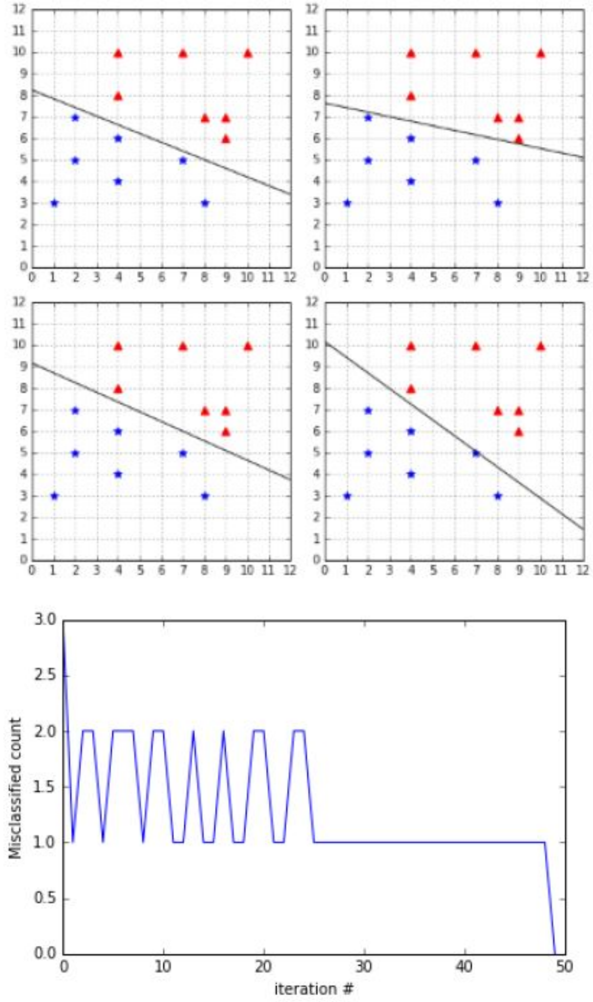
Perceptron Learning Algorithm

$$\operatorname{sign} f(x), \quad f(x) = x^\top w + b,$$

$$w_i = w_{i-1} + \gamma y_i x_i, \text{ if } y_i w^T x_i \leq 0$$

Iteratively separes linearly separable data:

- no generalization
- never ends if data are non-separable



What's the best separating hyperplane?

We need to compute a number that allows us to tell which hyperplane separates the data the best.

$$f = y(\mathbf{w} \cdot \mathbf{x} + b)$$

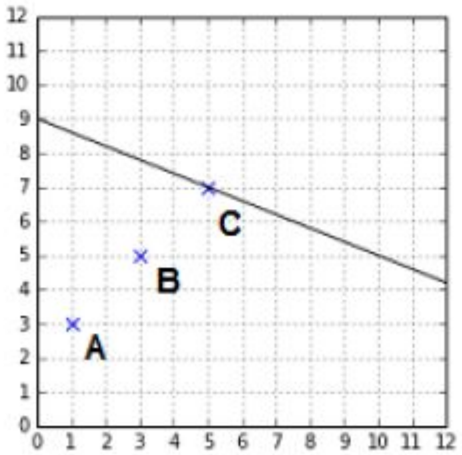
$$\gamma = y\left(\frac{\mathbf{w}}{||\mathbf{w}||} \cdot \mathbf{x} + \frac{b}{||\mathbf{w}||}\right)$$

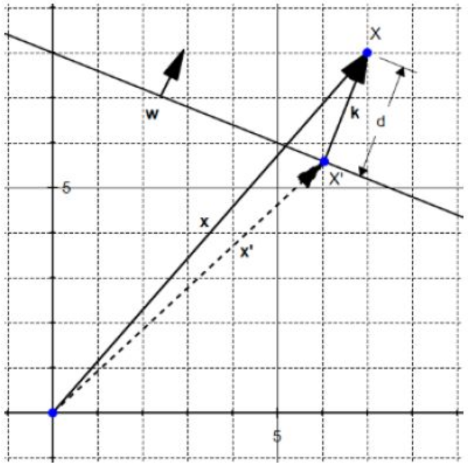
$$F = \min_{i=1...m} y_i(\mathbf{w} \cdot \mathbf{x}_i + b)$$

$$M = \min_{i=1...m} y_i\left(\frac{\mathbf{w}}{||\mathbf{w}||} \cdot \mathbf{x} + \frac{b}{||\mathbf{w}||}\right)$$

Functional Margin

Geometric Margin





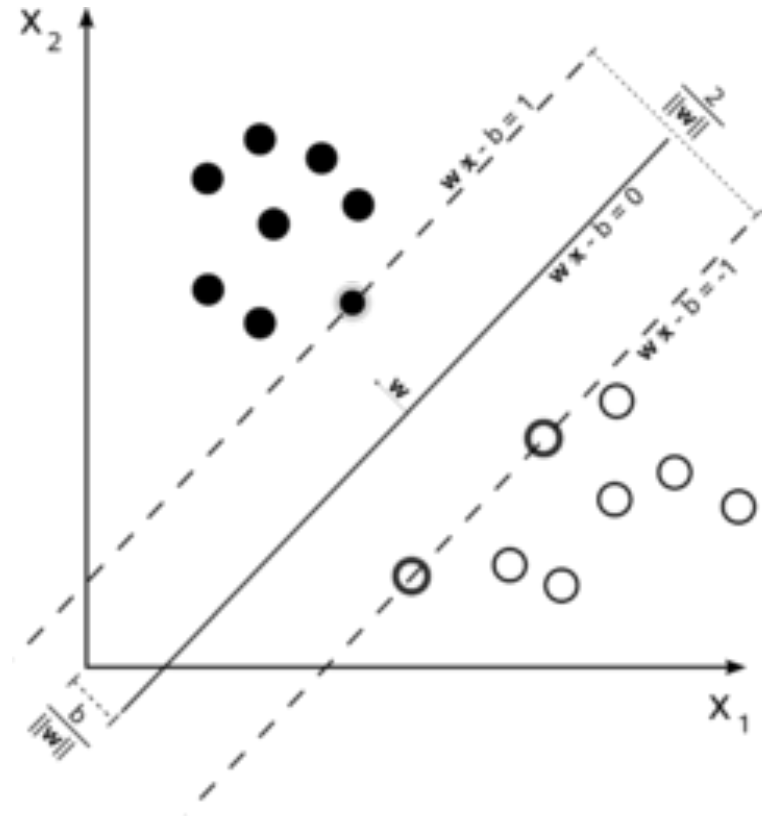
Support Vector Machines

SVM optimization problem *(hard margin)*:

$$\begin{aligned} &\underset{\mathbf{w},b}{\text{maximize}} && M \\ &\text{subject to} && \gamma_i \geq M, \quad i = 1, \ldots, m \end{aligned}$$

$$\begin{aligned} &\underset{\mathbf{w},b}{\text{maximize}} && \frac{F}{\|\mathbf{w}\|} \\ &\text{subject to} && f_i \geq 1, \quad i = 1, \ldots, m \end{aligned}$$

$$\begin{aligned} &\underset{\mathbf{w},b}{\text{minimize}} && \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to} && y_i(\mathbf{w} \cdot \mathbf{x}_i) + b \geq 1, \quad i = 1, \ldots, m \end{aligned}$$



Solution

- Lagrange multipliers
- Wolfe dual problem (Slater's condition)
- Karush-Kuhn-Tucker conditions

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{m} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1]$$

$$\nabla_{\mathbf{w}} \mathcal{L} = \mathbf{w} - \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i = \mathbf{0}$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i = 0$$

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{subject to} \quad & \alpha_i \geq 0, \text{ for any } i = 1, \ldots, m \\ & \sum_{i=1}^{m} \alpha_i y_i = 0 \end{aligned}$$

CVXOPT QP solver:

Computational complexity: $O(Dn^3)$

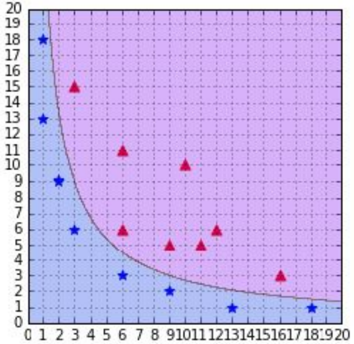
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{1}{2} x^T P x + q^T x \\ \text{subject to} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

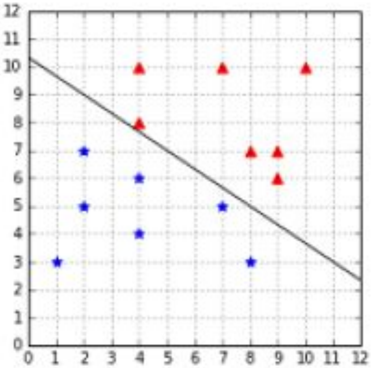
Not linearly separable data

Soft margin SVM

$$\begin{aligned} \underset{\mathbf{w},b,\zeta}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\zeta_i \\ \text{subject to} \quad & y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \zeta_i \\ & \zeta_i \geq 0 \quad \text{for any } i = 1,\ldots,m \end{aligned}$$

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \mathbf{x}_i\cdot\mathbf{x}_j \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, \text{ for any } i = 1,\ldots,m \\ & \sum_{i=1}^{m}\alpha_i y_i = 0 \end{aligned}$$





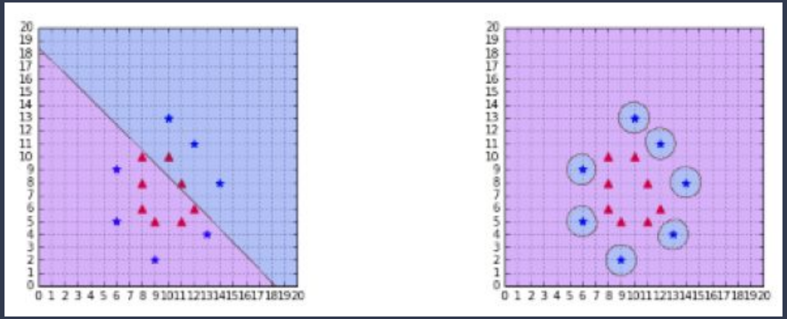
Kernel

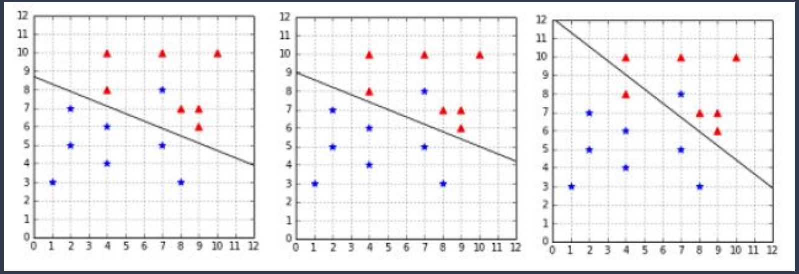
$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j) \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, \text{ for any } i = 1,\ldots,m \\ & \sum_{i=1}^{m}\alpha_i y_i = 0 \end{aligned}$$

Cross Validation

#X_train = 2000 for performance reasons

	linear	$\frac{1}{176+V(x)}$	$5\ e^{-5}$	$5\ e^{-4}$	1/176	$5\ e^{-2}$
0.1	0.87	0.88	0.49	0.89	0.87	0.49
1	0.87	**0.91**	0.87	0.91	0.9	0.63
10	0.86	0.9	0.9	0.9	0.9	0.64





Final Results

	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	0.0	0.91	0.84	0.89	0.89	0.92	0.92	0.93	0.83	0.9
car	0.91	0.0	0.94	0.92	0.95	0.95	0.93	0.94	0.89	0.84
bird	0.84	0.94	0.0	0.74	0.76	0.77	0.83	0.85	0.91	0.93
cat	0.89	0.92	0.74	0.0	0.78	0.68	0.82	0.83	0.92	0.9
deer	0.89	0.95	0.76	0.78	0.0	0.81	0.86	0.84	0.92	0.94
dog	0.92	0.95	0.77	0.68	0.81	0.0	0.83	0.83	0.95	0.93
frog	0.92	0.93	0.83	0.82	0.86	0.83	0.0	0.92	0.95	0.94
horse	0.93	0.94	0.85	0.83	0.84	0.83	0.92	0.0	0.94	0.91
ship	0.83	0.89	0.91	0.92	0.92	0.95	0.95	0.94	0.0	0.9
truck	0.9	0.84	0.93	0.9	0.94	0.93	0.94	0.91	0.9	0.0