

Компьютерное зрение '2014

Основы машинного обучения

Who? Александр Вахитов

When? December 16, 2014

Как применить машинное обучение?

- формирование обучающей выборки
- обучение детектора или классификатора
- тестирование на тестовой выборке

О чем мы поговорим

Общая идея

Основные
понятия

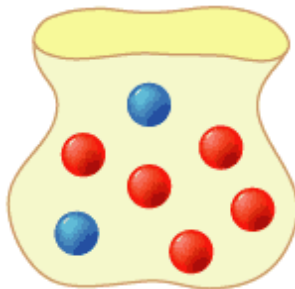
Практика
жизни

Модель
обучения

HOG-SVM для
распознавания
пешеходов

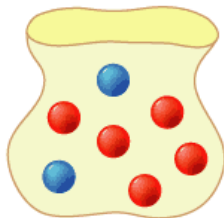
Эксперимент

Мешок с шарами 2 цветов



Достали N шаров, среди них ν - доля красных.
Задача: оценить, сколько красных в мешке?

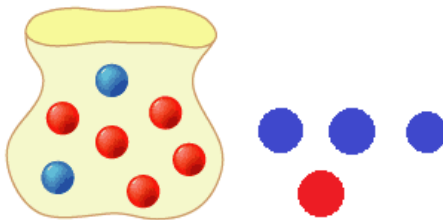
Эксперимент



$$P(\text{red}) = \mu; \quad P(\text{blue}) = 1 - \mu$$

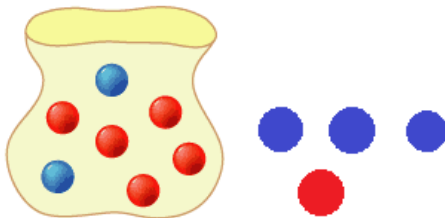
Необходимо найти μ

Эксперимент



Достали 4 шара из мешка. Что мы можем узнать о μ ?

Эксперимент



Общая интуиция:

Может быть, ν сильно отличается от μ

Скорее всего, ν примерно равно μ

Неравенство Бернштейна - Хефдинга

Неравенство Бернштейна (Хефдинга, Hoeffding)
Произведем выборку длины N , при этом все случайные величины, участвующие в выборке, независимы и одинаково распределены. Если значения с.в. ограничены, то справедливо неравенство:

$$P(|\nu - \mu| > \epsilon) \leq 2e^{-2\epsilon^2 N}$$

ν - среднее по выборке размера N

μ - математическое ожидание

Иными словами, $\mu = \nu$ асимптотически ($\nu \xrightarrow{N \rightarrow \infty} \mu$)

Обучение бинарного классификатора

- имеем выборку из N_p правильных объектов и N_n неправильных объектов
- разбиваем выборку на тренировочный и тестовый набор в пропорции 80-20
- выбираем дескриптор и алгоритм обучения
- строим ROC-кривую

Основные характеристики

- True Positive - истинно позитивные - истинные объекты, на которых сработал классификатор
- False Positive - ложно позитивные - ложные объекты, на которых сработал классификатор
- True Negative - истинно негативные - ложные объекты, на которых не сработал классификатор
- False Negative - ложно негативные - истинные объекты, на которых не сработал классификатор

ROC-кривая

Какова доля истинных объектов, на которых сработал классификатор, среди всех истинных объектов? (Кинул невод...)

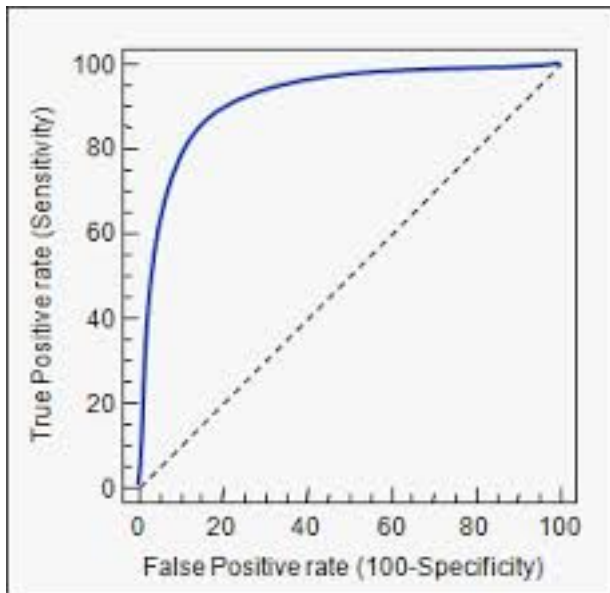
$$Recall = \frac{TP}{TP + FN}$$

Доля ложных срабатываний классификатора среди всех срабатываний классификатора:

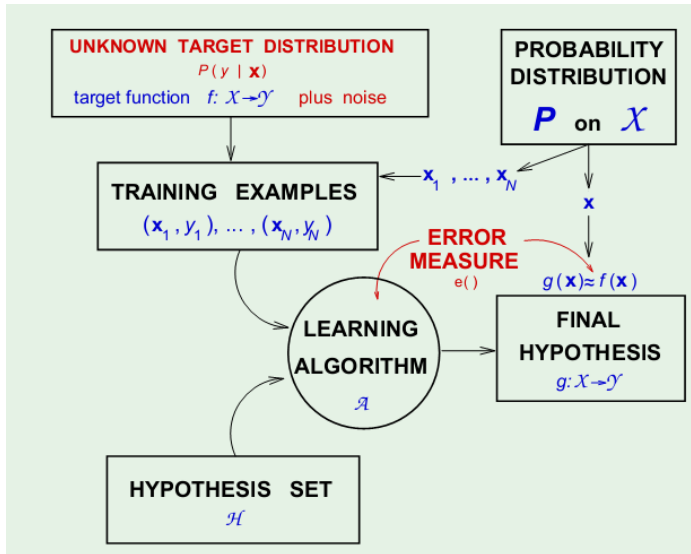
$$Fall - Out = \frac{FP}{FP + TP}$$

ROC-кривая (Receiver operating characteristic) - зависимость Recall от Fall-Out

ROC-кривая



Модель обучения (Y. Abu-Mostafa, Learning from Data)



Пояснения к модели обучения

- Пространство примеров X , распределение $P(X)$ на X задает частоту встречаемости отдельных примеров (напр., большинство людей на улице Петербурга - европеоиды)
- Неизвестная функция $f(x)$ определяет наблюдение для примера x , например x - конкретное лицо, $f(x)$ - его изображение
- Наблюдаем зашумленное значение $y = f(x) + v$, то есть задано распределение $P(Y|X)$
- Генерация выборки: генерируем x_1, \dots, x_N с помощью $P(X)$, затем генерируем y_i для каждого x_i с помощью $f(x)$ и $P(Y|X)$
- Используем обучающий алгоритм A , который призван выбрать финальную гипотезу $g(x)$ из множества гипотез H , так что $f(x) \approx g(x)$ для всей тренировочной выборки
- функция ошибки $e(f(x), g(x))$ задает штраф за неверные результаты при обучении

Пример 1. Распознавание пешеходов

(Dalal, Triggs, 2005)

Исходные данные



Пример 1. Детекция пешеходов. Дескриптор

(Dalal, Triggs, 2005)

- приводим к стандартному размеру
- делим на перекрывающиеся клетки и строим для них гистограммы градиентов по углу и длине
- формируем дескриптор из гистограмм

На полученных векторах обучаем линейный классификатор методом SVM:

$$g(x) = \sum_{i=1}^D w_i x_i + w_0,$$

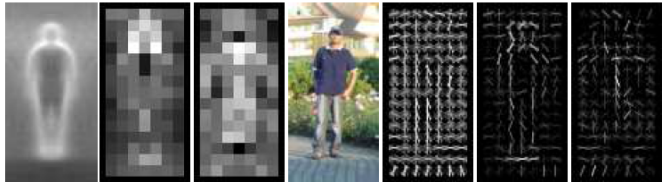
где D - длина дескриптора, w_i - веса, x_i - компоненты вектора дескриптора.

При $g(x) > 0$ - пешеход, иначе - нет.

Пример 1. Детекция пешеходов.

Дескриптор

(Dalal, Triggs, 2005)

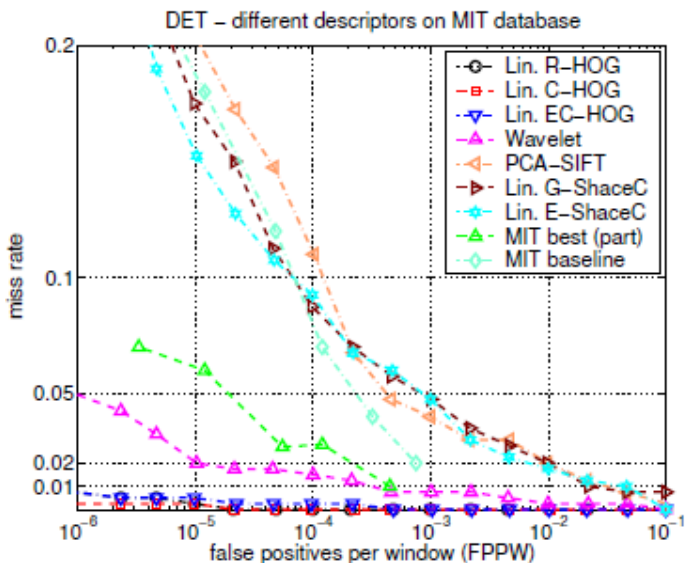


Пояснения к иллюстрации

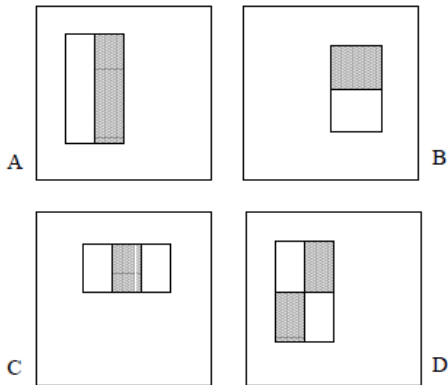
- слева - средний контур
- цвет клетки равен максимальному положительному весу для пикселей клетки
- цвет клетки равен максимальному модулю из отрицательных весов для пикселей клетки
- изображение
- его дескриптор
- дескриптор, домноженный на положительные веса
- дескриптор, домноженный на отрицательные веса

Пример 1. Детекция пешеходов

(Dalal, Triggs, 2005)



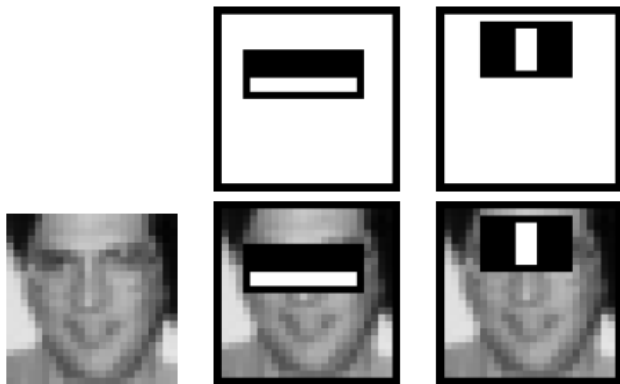
2. Haar - детекция лиц: базовые особенности (haar-like features)



складываем пиксели в окне: темный цвет с весом -1 ,
белый с весом $+1$

Как быстро посчитать такую величину?

2. Haar - детекция лиц: идея



Алгоритм AdaBoost

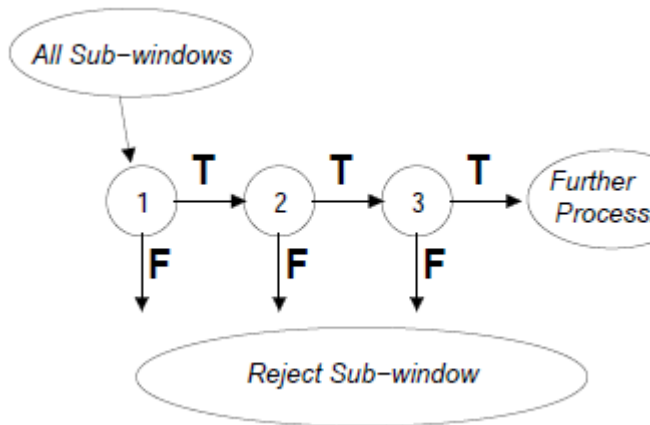
В цикле T раз:

- для каждой особенности j построить классификатор на ее основе h_j
- посчитать для каждого h_j ошибку классификации
- выбрать наилучший h_j , убрать особенность из множества доступных, добавить h_j с некоторым весом в множество избранных, обозначим его h_t

По избранным 'слабым' классификаторам h_j построить 'сильный'

$$h(x) = \begin{cases} 1, & \sum_{t=1}^T w_t h_t + w_0 \geq 0 \\ 0, & \sum_{t=1}^T w_t h_t + w_0 < 0 \end{cases}$$

Предлагается (Viola, Jones 2001) каскадный AdaBoost



Каскадный классификатор: стадия заканчивается, когда при условии $TP \leq r$ $FP \leq q$ (например: при $TP \leq 0.995$ $FP = 0.25$).

Сколько надо стадий в этом примере, чтобы общий FP стал меньше 0.01?