

Learning PAC

Andres J. Montenegro Bello*

GitHub : AndresJuniorMontenegro

Raul Huaman Pajares †

GitHub : rhuamanpa

Repositorio

<https://github.com/rhuamanpa/Pac-Learning>

ABSTRACT

We will analyze PAC learning, the type of uses that we can apply with PAC learning with machine learning and show some samples used by big companies like Facebook and Youtube to make incomes at the same time we will make a project using PAC learning using python that will recognize some facial gestures.

1 INTRODUCTION

La probabilidad aproximada correcta (PAC), propuesta por L. Valiant, es un marco de referencia estatico para aprender usar datos de entrenamiento.

En su forma mas simple y para un modelo tipico de tarea, la teoria del aprendizaje PAC intentan relacionar, precision y confianza estadistica del modelo al numero de ejemplos de entrenamientos usados.

Uno de los conceptos más importantes en este sentido es medir la complejidad de una clase de hipótesis H . En cualquier modelo de aprendizaje automático, el objetivo final es encontrar una clase de hipótesis que logre una alta precisión en el conjunto de entrenamiento y que tenga un bajo error de generalización en el conjunto de prueba. Para esto, requerimos que la clase de hipótesis H se aproxime al concepto de clase C que determina las etiquetas para la distribución D . Como tanto C como D son desconocidos, tratamos de modelar H en base al conjunto de muestras conocido S y sus etiquetas.

Generalizacion del error: El error de generalización de una hipótesis h es la expectativa del error en una muestra x elegida de la distribución D .

Error empirico: Esta es la media del error de la hipótesis h en la muestra S del tamaño m .

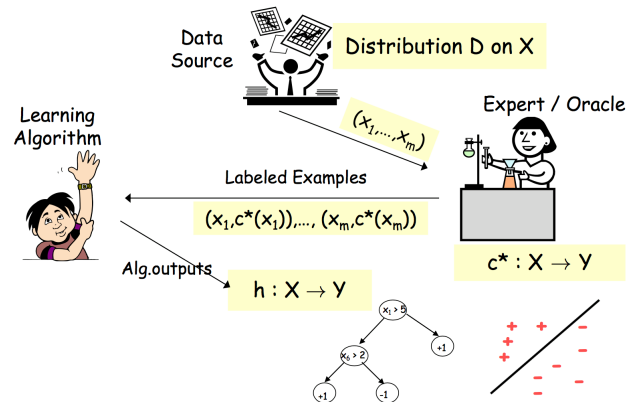
Habiendo definido el error de generalización y el error empírico de esta manera, podemos establecer el objetivo del aprendizaje de la siguiente manera.

El objetivo del aprendizaje es tener el error empírico aproximado al error de generalización con alta probabilidad.

Este tipo de marco de aprendizaje se conoce como Aprendizaje PAC (Probablemente Aproximadamente Correcto). Formalmente, una clase de concepto C es PAC – *aprendible* si hay algún algoritmo A para el cual el error de generalización en una muestra S derivado de la distribución D es muy bajo (menor que ϵ) con alta probabilidad (mayor que $1 - \delta$). En otras palabras, podemos decir que para una clase apta para PAC, la precisión es alta con buena confianza.

Se adjuntara una imagen en la cual dara una descripcion general del aprendizaje PAC.

PAC/SLT models for Supervised Learning



2 ESTADO DEL ARTE

El aprendizaje PAC(aprendizaje correcto probablemente aproximado) como su nombre lo dice generara probables situaciones para dar un resultado correcto aproximado, esta ciencia es una similitud o podemos decir una rama del machine learning y el deep learning, sus usos son muy dinamicos y eficaces, actualmente se utiliza como herramienta muy util veamos 3 ejemplos:

Identificación de temas: Clasificación multi-etiqueta de medio de artículos impresos deducido de una base de datos.

La data fue recolectada desde el Mayo del 2013 hasta Setiembre del 2013,

Los articulos fueron manualmente segmentados.

El texto de los articulos es representado usando "Una mochila de palabras modelo".

Esta base de datos cuenta con 301561 atributos numericos, 213 tipo de etiqueta y 99780 articulos, los cuales 64857 fueron data de entrenamiento y 34923 fueron data de prueba, el objetivo es predecir las etiquetas relevantes del conjunto de data de prueba.

Se adjunta el link del conjunto de data:

<https://www.kaggle.com/c/wise-2014/data>

Recomendación de películas: Sistema utilizado para predecir la afluencia que tendra una pelicula y su calificacion global, determinada por la calificacion previa a peliculas con similar caracteristica o etiqueta sacada de una base de datos ya sea de Netflix o Movie Lens

Se adjuntara los links de la base de datos de dichas plataformas de películas:

Netflix:

<https://www.netflixprize.com/leaderboard.html>

*e-mail: montenegroandresj@gmail.com

†e-mail: rahuamanpa@gmail.com

Movie Lens:

<https://grouplens.org/datasets/movielens/>

Video resumen: Selecciona semanticamente las partes más importantes de un video debido a una base de datos muy grande de metricas en la plataforma de *YOUTUBE*, esto es usado como grandes compañías como *Facebook* y *Youtube*, para que puedan insertar la publicidad en las partes mas relevantes del video subido y asi más probabilidad de que puedan ver la publicidad.

Se adjunta la data de *Youtube* y es :

<https://research.google.com/youtube8m/index.html>

3 DISEÑO DEL EXPERIMENTO

Se diseñara un experimento en la cual se capturara una imagen y este reconocera ciertos gestos faciales.

Debido a un conjunto de patrones establecidos.

Se implementara con el lenguaje de programación *Python*.