

Perceptron adaptativo

Diapositiva 1: El tema de la presentación es el perceptrón adaptativo

Diapositiva 2: El objetivo es aprender en ... ¿qué vamos a aprender?

Diapositiva 3: Semiespacios a partir de ejemplos tomados de forma uniforme en la esfera unitaria. Hablar no más de 20 sec acerca de la importancia de los semiespacios.

Diapositiva 5: En 1957, Frank Rosenblatt desarrollo el perceptrón a partir de la neurona de McCulloch Pitt. ¿Qué es un perceptrón?

Diapositiva 6: Un perceptrón es una unidad básica de inferencia en forma de discriminador lineal

Diapositiva 7: Explicar la animación. Entradas-> Pesos-> Sumatoria-> Función activadora(signo)-> Salida(semiespacio)

Diapositiva 8: Hipótesis: Función de R^n que cae en el conjunto $-1,1$

Dataset: Pares de $x, f^*(x)$, explicar que es el f^* y de donde tomamos los x

Diapositiva 9: Definiciones un poco más formales de lo anterior.

Diapositiva 10: Explicación del error(animación), hablar de la esfera y esperar a que acabe la animación.

Diapositiva 11: Explicar nueva definición del error a partir del ángulo y el producto punto.

Diapositiva 12: Una actualización luce: Explicar que es el η (ratio de aprendizaje).

Diapositiva 13: Error después de una actualización.

Diapositiva 14: Explicar la minimización. Queremos minimizar el error en función de η , que es lo importante de este algoritmo. Arccos es decreciente, solo falta maximizar.

Diapositiva 15: Derivando e igualando a 0. Explicar sustitución d y d^* .

Diapositiva 16: Despejamos el ratio de aprendizaje y obtenemos la ecuación ... Recordar que ese producto punto es el cos de β

Diapositiva 17: Luego el ratio de aprendizaje optimo es dado por... Esto no es practico ya que necesitamos conocer valores de la función objetivo en cada paso.

Diapositiva 18: Solución. Suponer que el error de w es pequeño, el ángulo, coseno también y los valores esperados también. $d^*=d$

Diapositiva 19: De aquí sale la idea de tomar el ratio de aprendizaje como ...

Diapositiva 20: Luego el algoritmo es: Entrada: número de ejemplos de forma uniforme en la esfera unitaria.

Salida: Hipótesis w

Explicar el resto de pasos tal como esta y hablar de la definición de contraejemplo en el si

Diapositiva 21: Algunas propiedades son: Norma unitaria y error monótono decreciente. Animación (aparece gráfica) comparación entre diferentes versiones del algoritmo del perceptrón. Como se puede ver, la más cercana a la óptima es el adaptativo.

Diapositiva 22: Pasar a los lemas. Coseno esperado del ángulo de la hipótesis... Producto esperado de las distancias ... Cota inferior de Chernoff(enlazar con lo que ya vimos en clase).

Diapositiva 23: Ahora sí, demostremos el teorema. Leer enunciado del teorema

Diapositiva 24: La demostración es: tomamos el w_k que es la hipótesis k -ésima que genera el algoritmo, β_k es el ángulo. El valor esperado del error se define como:...

(esperar)

Por desigualdad de Jensen tenemos:

(esperar)

Por propiedades del arccos

(esperar)

Por lema Valor esperado del coseno. Pasamos a términos exponenciales

(esperar)

Por propiedades básicas de la exponencial

(esperar)

Después de al menos k iteraciones, donde k es tres n logaritmo natural de 2 sobre ϵ por δ contraejemplos.

Suponga que tenemos un error mayor a ϵ . Luego, tomando s como ... esperamos tener al menos $2k_0$ contraejemplos.

(esperar)

Para ver esto podemos usar la cota inferior de Chernoff(probabilidad del número de contraejemplos menor a k_0)

(esperar)

Si el error de la hipótesis w_k es mayor que ϵ

Hay menos de k_0 contraejemplos o el valor de la variable aleatoria excede su valor esperado en un múltiplo de 2 sobre δ .

(esperar)

Mencionar cota de uniones y desigualdad de Markov (las vimos en clase, mencionar eso). Luego la probabilidad de que la hipótesis tenga un error mayor a ϵ es a lo más δ .

Tomando el complemento de la probabilidad. Queda demostrado

Diapositiva 25: Conclusiones y problemas abiertos

Diapositiva 26: Conclusiones: El algoritmo es eficiente (converge en “pocos” pasos y con pocos ejemplos) en la esfera unitaria tomando de forma uniforme los ejemplos

Diapositiva 27: Problemas abiertos: ¿En qué otras distribuciones de probabilidad es eficiente?

Diapositiva 28: Gracias por la atención. Fin.