

Máquinas de Boltzmann: Ecuaciones, Gradientes y Aplicaciones

December 10, 2024

Introducción a las Máquinas de Boltzmann

¿Qué son las Máquinas de Boltzmann?

Las Máquinas de Boltzmann son un modelo estocástico de redes neuronales que utilizan una función de energía para modelar dependencias probabilísticas entre las neuronas.

- ▶ **Modelo probabilístico:** Las variables visibles y ocultas están conectadas a través de una función de energía.
- ▶ **Entrenamiento:** Minimización de una función de costo, generalmente usando el algoritmo de Contraste de Divergencia (CD).
- ▶ **Aplicaciones:** Son útiles para entender redes neuronales profundas y aprender representaciones latentes de los datos.

Ecuaciones Fundamentales

La función de costo que se optimiza en las Máquinas de Boltzmann es:

$$L(\theta, D) = \frac{1}{N} \sum_{x^i \in D} \log p(x^i) \quad (1)$$

donde:

- ▶ D es el conjunto de datos.
- ▶ $p(x^i)$ es la probabilidad de un dato específico x^i .

La función de costo negativa, que se minimiza, es:

$$l(\theta, D) = -L(\theta, D) \quad (2)$$

Probabilidad en las Máquinas de Boltzmann

La probabilidad de una configuración visible x se calcula sumando sobre todas las posibles configuraciones de las variables ocultas h :

$$p(x) = \sum_h \frac{e^{-E(x,h)}}{Z} \quad (3)$$

donde:

- ▶ $E(x, h)$ es la energía de la configuración conjunta de x y h .
- ▶ Z es la función de partición, que normaliza la probabilidad.

Energía de la Configuración

La energía de una configuración de las variables visibles x y las ocultas h se expresa como:

$$E(x, h) = -b'x - c'h - h'Wx \quad (4)$$

donde:

- ▶ b es el vector de sesgos de las variables visibles.
- ▶ c es el vector de sesgos de las variables ocultas.
- ▶ W es la matriz de pesos entre las variables visibles y ocultas.

Gradientes de la Función de Costo

Los gradientes de la log-verosimilitud con respecto a los parámetros de la red (pesos y sesgos) se calculan de la siguiente forma:

Para la matriz de pesos W_{ij} :

$$-\frac{\partial \log p(x)}{\partial W_{ij}} = E_x [p(h_i|x) \cdot x_j] - x_j^i \cdot \sigma(W_i \cdot x^i + c_i) \quad (5)$$

Para los sesgos de las variables ocultas c_i :

$$-\frac{\partial \log p(x)}{\partial c_i} = E_x [p(h_i|x)] - \sigma(W_i \cdot x^i) \quad (6)$$

Para los sesgos de las variables visibles b_j :

$$-\frac{\partial \log p(x)}{\partial b_j} = E_x [p(x_j|h)] - x_j^i \quad (7)$$

donde σ es la función sigmoide, que activa las neuronas.

Aplicaciones de las Máquinas de Boltzmann Restringidas (RBM)

Las RBM tienen una amplia gama de aplicaciones en diversos campos de la ciencia y la industria:

- ▶ **Reconocimiento de patrones:** Se utilizan para extracción de características en problemas de reconocimiento de patrones, como la interpretación de texto manuscrito o patrones aleatorios.
- ▶ **Motores de recomendación:** Se emplean en técnicas de filtrado colaborativo para predecir qué recomendar al usuario, como en recomendaciones de películas o libros.
- ▶ **Reconocimiento de objetivos en radar:** En sistemas de radar con baja relación señal/ruido, las RBM ayudan a detectar el intrapulso.

Características de las Máquinas de Boltzmann Restringidas

Las RBM presentan varias características clave que las hacen únicas y eficaces en el aprendizaje no supervisado:

- ▶ **Estructura recurrente y simétrica:** Las RBM utilizan una estructura recurrente en la que las unidades visibles y ocultas están conectadas de manera simétrica.
- ▶ **Aprendizaje no supervisado:** Son algoritmos de aprendizaje no supervisado, es decir, no requieren respuestas etiquetadas.
- ▶ **Asociación de alta probabilidad a estados de baja energía:** En su proceso de aprendizaje, las RBM intentan asociar alta probabilidad a los estados de baja energía.
- ▶ **Sin conexiones intra-capa:** No existen conexiones dentro de las capas visibles o ocultas.

Entrenamiento por Contraste de Divergencia

El entrenamiento de una RBM se realiza mediante el algoritmo de Contraste de Divergencia (CD), que aproxima el gradiente de la log-verosimilitud. Los pasos son:

1. Inicializar el modelo con los datos visibles.
2. Actualizar las variables ocultas usando una probabilidad condicionada.
3. Generar una nueva configuración visible condicionada por las variables ocultas.
4. Calcular la diferencia entre las configuraciones visibles antes y después de la actualización y ajustar los pesos.

Este proceso se repite iterativamente, permitiendo la optimización eficiente del modelo.

Conclusión

Las Máquinas de Boltzmann son un modelo potente para aprender representaciones latentes de los datos. El algoritmo de Contraste de Divergencia permite optimizar los parámetros del modelo, aunque su cálculo puede ser costoso. Las variantes como las Máquinas de Boltzmann Restringidas (RBM) hacen más eficiente este proceso.

Desafíos

A pesar de su potencia, las Máquinas de Boltzmann sufren de problemas computacionales debido al cálculo de los gradientes, lo que ha motivado el desarrollo de métodos alternativos como las Redes Neuronales Convolucionales y otras variantes.