

1.2 Cross Entropy Multilabel

Summary

Lo único que hace falta saber.....	1
Bibliography.....	1

Lo único que hace falta saber

Se utiliza en problemas de clasificación de múltiples clases en las que cada ejemplo puede pertenecer a ninguna, una o varias clases, no como en el caso de softmax + cross entropy, que cada ejemplo debe estrictamente pertenecer a una clase

Para empezar, la función de activación de la capa de salida debe ser Sigmoid, el resto no importa mucho mientras no sea lineal

La función de pérdida para cada nodo de salida se expresa como (siendo la y sin gorrita la salida esperada, y la y con gorrita la salida predicha por la red):

$$\mathcal{L}(\theta) = -[(1 - y) \log (1 - \hat{y}) + y \log \hat{y}]$$

Y la derivada de la función de coste con respecto al valor de salida de cada neurona en la capa de salida es:

$$\begin{aligned}\frac{\partial \mathcal{L}(\theta)}{\partial \hat{y}} &= \frac{\partial}{\partial \hat{y}} \{-(1 - y) \log (1 - \hat{y}) - y \log \hat{y}\} \\ &= (-)(-1) \frac{(1-y)}{(1-\hat{y})} - \frac{y}{\hat{y}} \\ &= \frac{\hat{y}(1-y) - y(1-\hat{y})}{(1-\hat{y})\hat{y}} \\ &= \frac{\hat{y} - y}{(1-\hat{y})\hat{y}}\end{aligned}$$

Bibliography

Página con las expresiones: <https://prvnk10.medium.com/sigmoid-neuron-and-cross-entropy-962e7ad090d1>