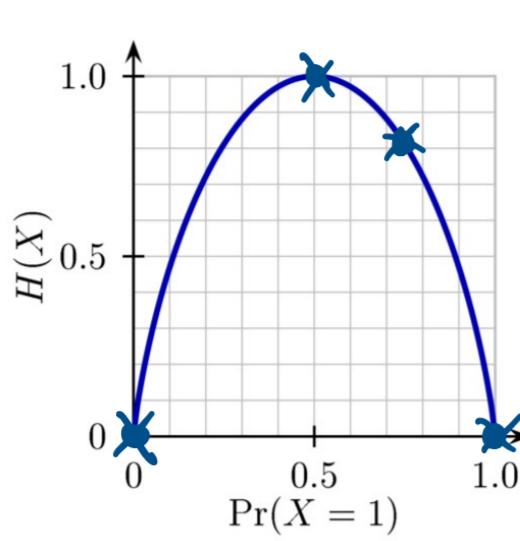


FUNZIONE COSTO

$$J(\alpha) = \frac{1}{m} \sum_{i=1}^m \text{Loss}(h_\alpha(x^{(i)}), y^{(i)}) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log(h_\alpha(x^{(i)})) + (1-y^{(i)}) \log(1-h_\alpha(x^{(i)})) \right]$$

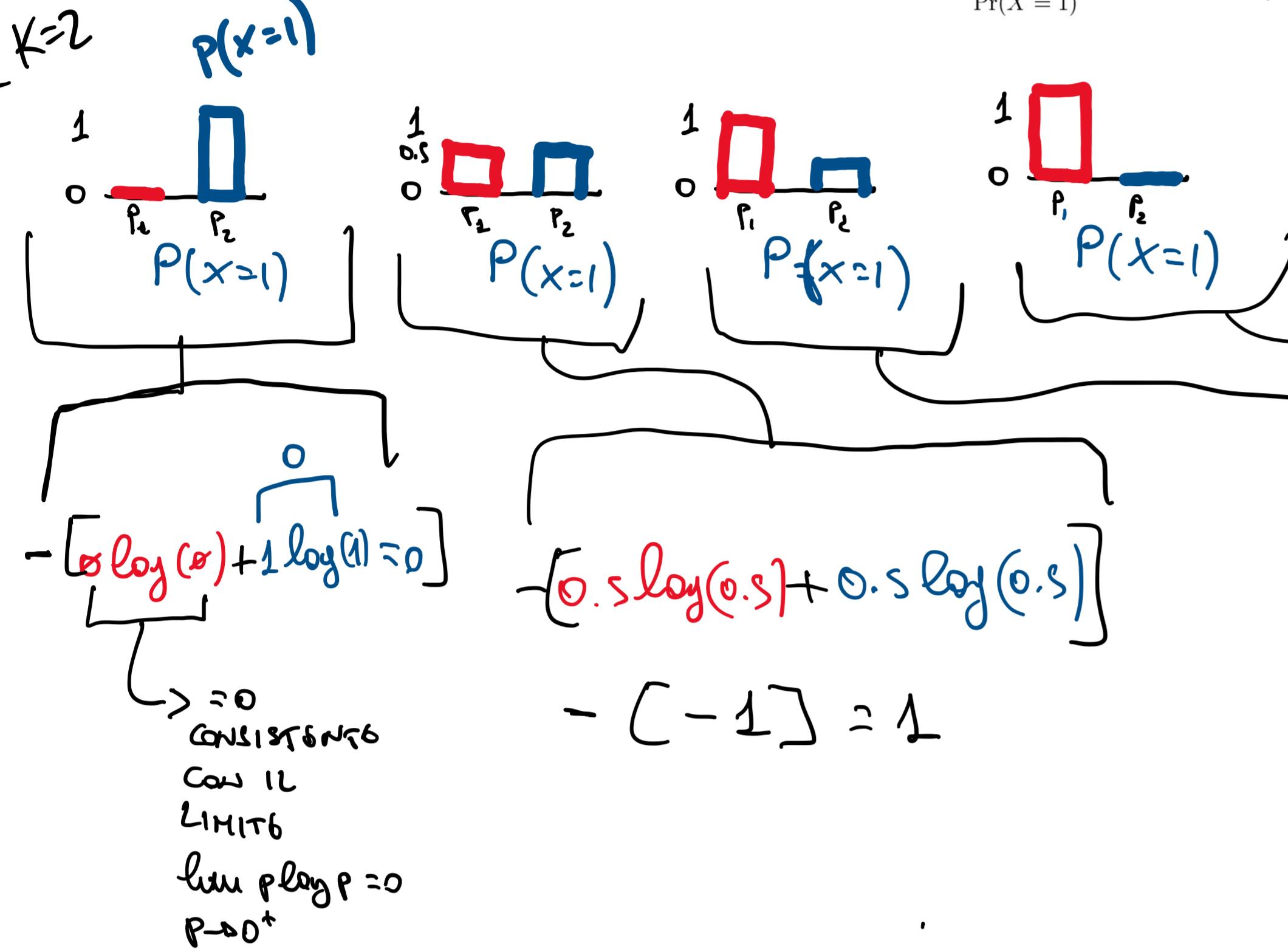
ENTROPY

$$H(X) = - \sum_{k=1}^K p_k \log(p_k)$$

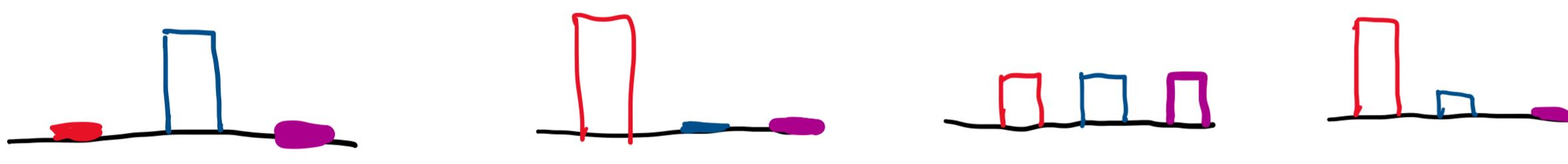


- MISURA
 - INEGALITÀ
 - MISURA DEL "DISORDINE" TRA CLASSI (DISTRIBUZIONE NELLA MISTRAZIONE)
 - QUANTO LA PROBABILITÀ È CONCENTRATA SU UNA CLASSE
 - INFORMAZIONE (IN QUESTO CASO, LA DISTRIBUZIONE NON È INFORMATIVA SE NON SO DECIDERE)

PER K=2



IL RAGIONAMENTO VALGONO PER PIÙ CLASSI



ESERCIZIO: ORDINA LE DISTRIBUZIONI IN ORDINE CRESCENTE RISPETTO L'ENTROPIA

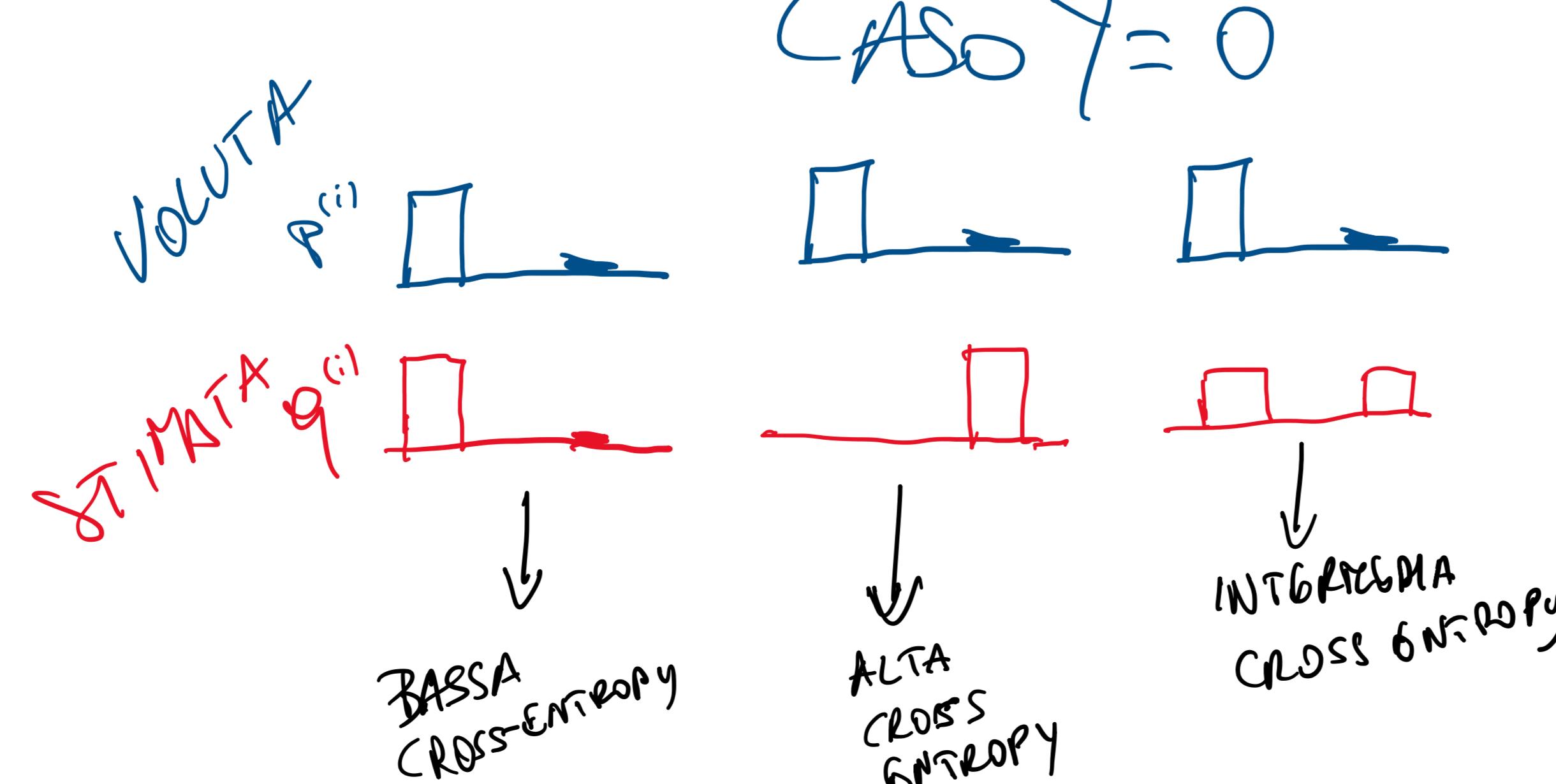
CROSS-ENTROPY (LOSS)

$$H(x) = \frac{1}{m} \sum_{i=1}^m \left(- \sum_{k=1}^K p_k^{(i)} \log(q_k^{(i)}) \right)$$

DISTRIBUZIONE DATA DAL CLASSIFICATORE

DISTRIBUZIONE IDEALE = QUELLA CHE VOGLIAMO OTTENERE. CORRISPONDE ALLA NOSSA $y^{(i)}$

MISURA LA "DISTANZA"
TRA DUE DISTRIBUZIONI DI PROBABILITÀ P E Q
 P È LA PROBABILITÀ REALE (GROUND TRUTH)
 q È LA PROBABILITÀ STIMATA
SI CONSIDERANO LE DISTRIBUZIONI SU TUTTI I PATTERN
→ VOGLIAMO CAMBIARE q
TOSI CHE IL PROB. DI PROBABILITÀ PISUTTO IN TUTTI I PATTERN IN INPUT (MINIMIZZARLO)



CASE $y^{(i)} = 1$

