

# Machine Learning

Soft-Max

---

Exemple sur une Sortie Softmax :

$$\forall x_i \in C, S(x_i) = \frac{e^{x_i}}{\sum_{x_j \in C} e^{x_j}}$$

Softmax  $\Rightarrow$  loss = cross-entropie + L2(W)

loss :

$$L(y^*, y) = H(y^*, y) + \frac{1}{2} \lambda \sum_{i \in P} w_i^2$$

$$H(y^*, y) = - \sum_{(i \in I)} y_i * \ln(y_i)$$

gradient :

$$\frac{\partial L(x_i)}{\partial w_i} = (S(x_i) - \delta_{i=i^*})X + \lambda W_i$$

Learning rate évolutif et adapté :

- SGD :  $grad = \alpha * grad_{t-1}(x) + \beta * grad_t$  où  $\alpha$  et  $\beta$  sont positifs et de somme 1.
- AdaGrad : gradient adaptatif par pondération des mises à jour
- Adam : pondération fine et adaptative des mises à jour de chaque poids

Intégrer des contraintes dans le gradient par la loss. On donne une direction différente à l'objectif et donc une surface des paramètres en fonction de l'erreur différente.

$$grad = grad_x(x) + grad_{contrainte1} + grad_{contrainte2} + \dots$$

- ajout de la norme L2 des poids à la loss => contraint les poids à rester faibles
- softmax => cross-entropie => sparsification des poids





ML WEEK

- 
- 
- 
- 
- 
- 
- 
-