

• Multiclass Logistic Regression:

$$P(y = \hat{y} | \bar{x}) = \frac{e^{\bar{w}^T f(\bar{x}, \hat{y})}}{\sum_{y \in \mathcal{Y}} e^{\bar{w}^T f(\bar{x}, y)}}$$

↑
must sum
to 1

← DF

Generalization

in binary LR:

$$y = +1 : e^{\bar{w}^T f(\bar{x})}$$

$$y = -1 : e^{\bar{w}^T \bar{s}} = 1$$

$$\frac{\partial}{\partial \bar{w}} \text{loss}(\bar{x}^{(i)}, y^{(i)}, \bar{w}) = -f(\bar{x}, y^{(i)}) + \sum_{y' \in \mathcal{Y}} P(y' | \bar{x}) f(\bar{x}, y')$$

• $P(y^{(i)} | \bar{x}) \approx 1 : -f(\bar{x}, y^{(i)}) + f(\bar{x}, y^{(i)}) \approx 0$

• $P(y_{\text{true}} | \bar{x}) \approx 1 : -f(\bar{x}, y^{(i)}) + f(\bar{x}, y_{\text{true}})$

SGD: $-\alpha \cdot \frac{\partial}{\partial \bar{w}}$