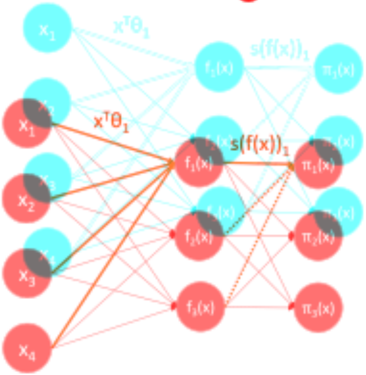
Introduction to Machine Learning

Multiclass Classification

Softmax Regression



Learning goals

- Know softmax regression
- Understand that softmax regression is a generalization of logistic regression

Learning goals
Know softmax regression
Understand that softmax regression is a generalization of logistic regression

... TO SOFTMAX REGRESSION / 2

- The softmax function is a generalization of the logistic function. For $g = 2$, the logistic function and the softmax function are equivalent.
- Instead of the **Bernoulli** loss, we use the multiclass **logarithmic loss**

$$L(y, \pi(\mathbf{x})) = -\sum_{k=1}^{g} \mathbb{1}_{\{y=k\}} \log\left(\pi_k(\mathbf{x})\right).$$

- Note that the softmax function is a "smooth" approximation of the arg max operation, so $s((1, 1000, 2)^T) \approx (0, 1, 0)^T$ (picks out 2nd element!).
- Furthermore, it is invariant to constant offsets in the input:

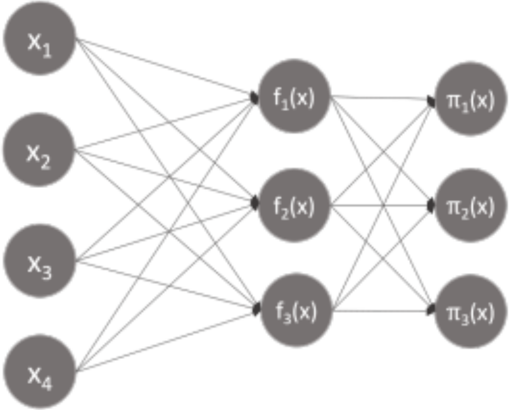
$$s(f(\mathbf{x})+\mathbf{c}) = \frac{\exp(\boldsymbol{\theta}_k^\top \mathbf{x} + c)}{\sum_{j=1}^{g} \exp(\boldsymbol{\theta}_j^\top \mathbf{x} + c)} = \frac{\exp(\boldsymbol{\theta}_k^\top \mathbf{x}) \cdot \exp(c)}{\sum_{j=1}^{g} \exp(\boldsymbol{\theta}_j^\top \mathbf{x}) \cdot \exp(c)} = s(f(\mathbf{x}))$$

LOGISTIC VS. SOFTMAX REGRESSION

	Logistic Regression	Softmax Regression
$\mathcal{Y}$	$\{0, 1\}$	$\{1, 2, ..., g\}$
Discriminant fun.	$f(\mathbf{x}) = \boldsymbol{\theta}^\top \mathbf{x}$	$f_k(\mathbf{x}) = \boldsymbol{\theta}_k^\top \mathbf{x}, k = 1, 2, ..., g$
Probabilities	$\pi(\mathbf{x}) = \frac{1}{1+\exp\left(-\boldsymbol{\theta}^\top \mathbf{x}\right)}$	$\pi_k(\mathbf{x}) = \frac{\exp(\boldsymbol{\theta}_k^\top \mathbf{x})}{\sum_{j=1}^g \exp(\boldsymbol{\theta}_j^\top \mathbf{x})}$
$L(y, \pi(\mathbf{x}))$	Bernoulli / logarithmic loss $-y \log\left(\pi(\mathbf{x})\right) - (1-y)\log\left(1-\pi(\mathbf{x})\right)$	Multiclass logarithmic loss $-\sum_{k=1}^g [y = k] \log\left(\pi_k(\mathbf{x})\right)$

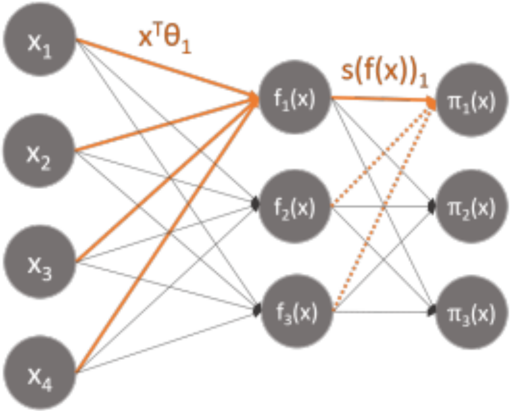
LOGISTIC VS. SOFTMAX REGRESSION

We can schematically depict softmax regression as follows:



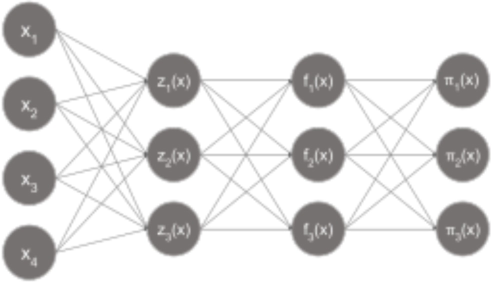
LOGISTIC VS. SOFTMAX REGRESSION

We can schematically depict softmax regression as follows:



GENERALIZING SOFTMAX REGRESSION / 2

For example for a **neural network** (note that softmax regression is also a neural network with no hidden layers):



Remark: For more details about neural networks please refer to the lecture **Deep Learning**.