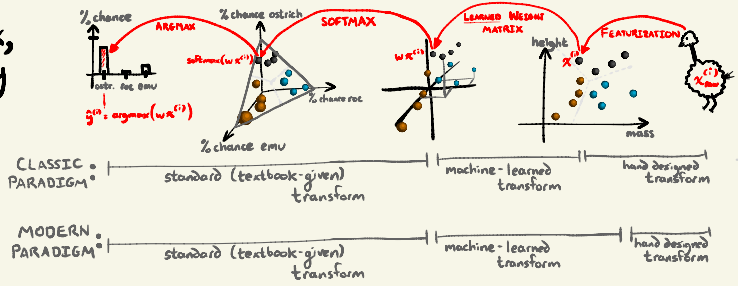
Softmax @ 5 Levels of Sophistication

Sam's Explainers
2022-06-23

0. softmax, visually
A. lay person
B. MFBT developer
C. deep architect
D. FBR developer
E. researcher

0. softmax, visually



A. lay person

Softmax is... how we pick 'the' best answer while being fair about ties

enables ML models to say "I don't know"

- humility allows step-by-step learning
- at deployment: calibrated confidences

B. MFBT developer

Softmax ... normalizes real-valued scores to probabilities

$$\text{softmax}: \mathbb{R}^K \to \{(p_i : 0 \le i < k) \in \mathbb{R}^K : (\forall 0 \le i < k : 0 \le p_i) \;\&\; (\textstyle\sum p_i = 1)\} \subseteq \mathbb{R}^K$$

$$(z_i : 0 \le i < k) \longmapsto \left(\frac{\exp(z_i)}{\sum_j \exp(z_j)} : 0 \le i < k\right)$$

```
z = np.matmul(w, x[i])
e = np.exp(z)
n = np.sum(e)
p = e/n
```

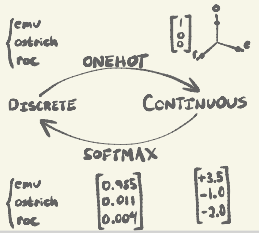
one-hot

```
dl_dp = (-1./p) * y[i]
dl_dn = -np.dot(dl_dp, e)/n**2
dl_de = dl_dp/n + dl_dn
dl_dz = dl_de * e
dl_dw = np.outer(dl_dz, x[i])
```

`dl_dz = p - y[i]`

C. deep architect

Softmax... allows continuous methods to solve discrete tasks



D. FBR developer

Softmax is... a convenient interface w/ likelihood loss

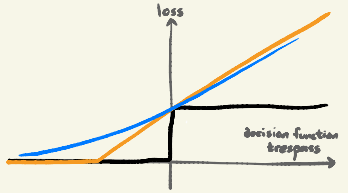
convex, so easy to optimize (cf. 01)

tight surrogate for 0-1 loss

dec. func. trespass $= \left(-\underbrace{y^{(i)}\ \vec{w} \cdot \vec{x}^{(i)}}\right)$

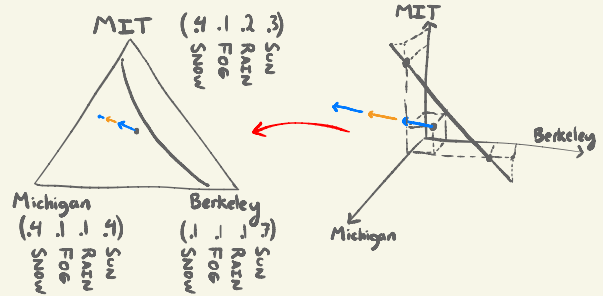
hinge loss $= \max(0, 1 + \text{DFT})$

softplus loss $= \log(1 + \exp(\text{DFT})) = \log\left(\frac{\exp(-\text{DFT})}{1 + \exp(-\text{DFT})}\right) = \left(\log \text{softmax} \begin{bmatrix} [\text{C-DFT}] \\ \text{C} \end{bmatrix}\right)[0]$



E. researcher

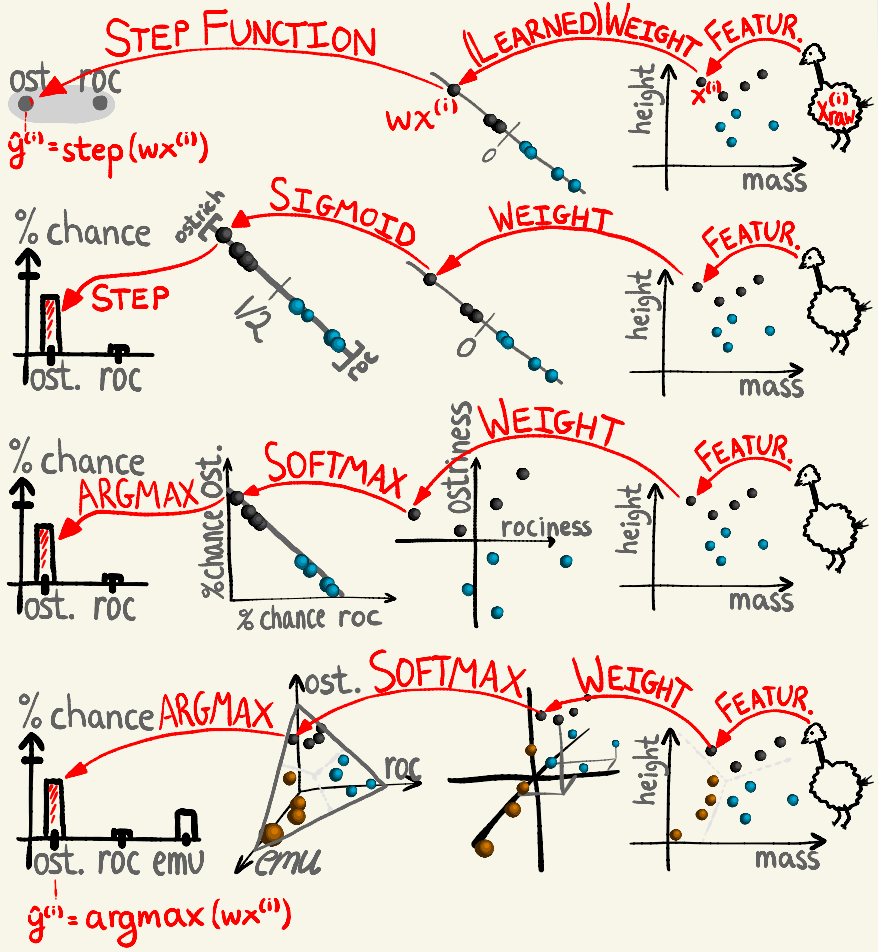
Softmax is... the canonical e-flat parameterization of the simplex



MIT	Mich	Berk
(.33	.33	.33)
observe SNOW ×(.4 .4 .1)		
(.13	.13	.03)
(.44	.44	.11)
observe SNOW ×(.4 .4 .1)		
(.17	.17	.01)
(.48	.48	.03)
observe SNOW ×(.4 .4 .1)		
(.19	.19	.003)
(.496	.496	.008)

SOFTMAX @ 5 LEVELS OF SOPHISTICATION

SAM'S EXPLAINERS
2022-06-23



0. softmax, visually
A. lay person
B. MFBT developer
C. deep architect
D. FBR developer
E. researcher

SOFTMAX is... how we pick 'the' best answer while being fair about ties.

SOFTMAX enables ML models to say "I don't know"

- humility allows step-by-step learning
- at deployment: calibrated confidences