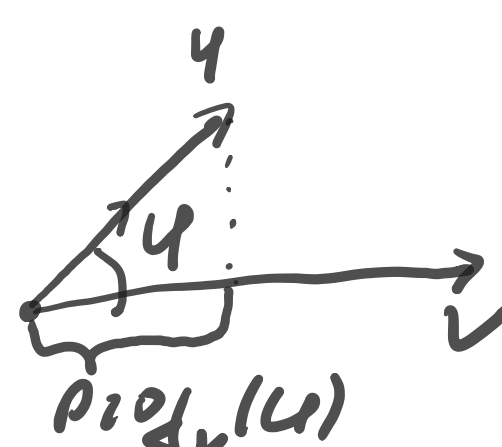


• Math refresher

$$\langle u, v \rangle = u_1 \cdot v_1 + u_2 \cdot v_2$$

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$



$$\begin{cases} \langle u, v \rangle = |v| \text{proj}_v(u) = |v||u| \cos \varphi \\ \text{proj}_v(u) = \cos \varphi |u| \end{cases}$$

$$\cos \varphi = \frac{\langle u, v \rangle}{|u||v|}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad Av = v_1 \begin{bmatrix} 1 \\ c_1 \end{bmatrix} + v_2 \begin{bmatrix} 1 \\ c_2 \end{bmatrix}$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \Rightarrow \tilde{I}v = \begin{bmatrix} c_1 & c_2 \\ b & d \end{bmatrix}$$

$$Av = \begin{bmatrix} av_1 + bv_2 \\ cv_1 + dv_2 \end{bmatrix}$$

$$A = \begin{bmatrix} -z_1 & - \\ -z_2 & - \end{bmatrix} \Rightarrow Av = \begin{bmatrix} \langle z_1, v \rangle \\ \langle z_2, v \rangle \end{bmatrix}$$

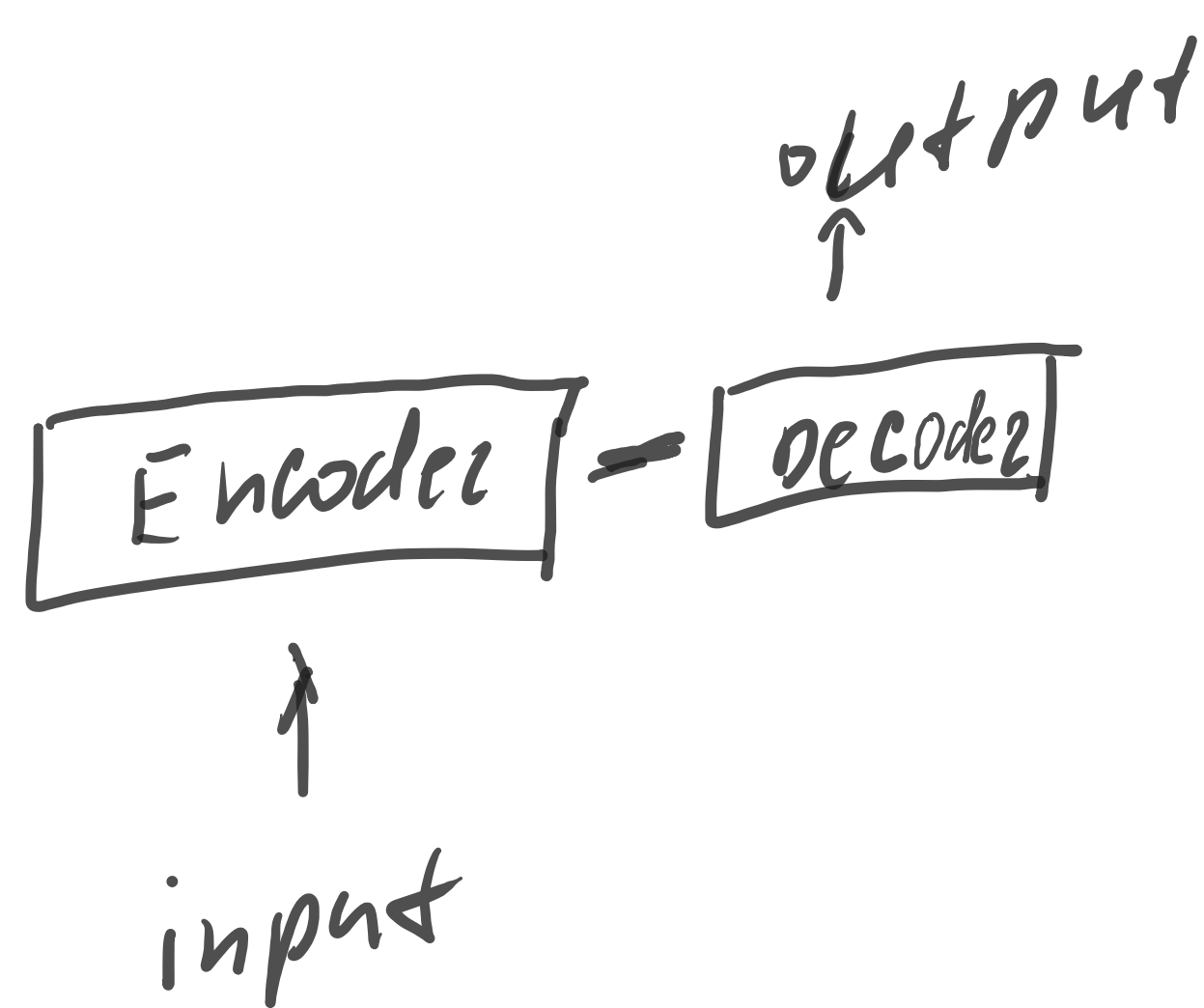
• Temperature scaling

$$p(i) = \frac{\exp(a_i)^{\frac{1}{T}}}{\sum_{j=1}^V \exp(a_j)^{\frac{1}{T}}}$$

$$\text{softmax}[a_1, a_2, \dots, a_V]$$

$$\exp(\log(a^t)) \Rightarrow a^t \Rightarrow t \log(a) \Rightarrow \exp(t \log(a))$$

$$p(i) = \frac{\exp(a_i / T)}{\sum_{j=1}^V \exp(a_j / T)}$$



• BPE

abacaba

[a, b, cab, aba, abac, abacaba]

0. ab - 2 1. aba - 2 2. abac - 1
 ba - 2 ac - 1 cab - 1
 ac - 1 cab - 1

• BBPE

unicode

utf8

A = byte1 byte2