22.4.2020 ERM for student-t likelihood

$$\operatorname*{argmax}_w \hat{p}(y_{1:n} | x_{1:n}, w)$$

$$= \operatorname*{argmin}_w - \sum_{i=1}^{n} \log \hat{p}(y_i | x_i, w)$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu\sigma^2}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(y_i - w^T x_i)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

Normalization constant $Z^{-1}(\nu, \sigma^2)$

$$= \operatorname*{argmin}_w - \sum_{i=1}^{n} \log Z^{-1}(\nu, \sigma^2) \left(1 + \frac{(y_i - w^T x_i)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

$$= \operatorname*{argmin}_w -n \cdot \log Z^{-1}(\nu, \sigma^2) + \sum_{i=1}^{n} \left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{(y_i - w^T x_i)^2}{\nu\sigma^2}\right)$$

$$= \operatorname*{argmin}_w \sum_{i=1}^{n} \log\left(1 + \frac{(y_i - w^T x_i)^2}{\nu\sigma^2}\right)$$

$$= \operatorname*{argmin}_w \sum_{i=1}^{n} \underbrace{\log\left(1 + \frac{(y_i - w^T x_i)^2}{c}\right)}_{\ell_{\text{Student-t}}(w; y_i, x_i)} \quad \text{for } c = \nu\cdot\sigma^2$$

$$\ell_{\text{Student-t}}(w; y_i, x_i) \leq \frac{(y_i - w^T x_i)^2}{c} = \frac{1}{c} \cdot \ell_2(w; y_i, x_i)$$

since $\log(1+x) \leq x$

$\log(1+x) \simeq x$ for $x \simeq 0$; $\log(1+x) \ll x$ for $x \gg 0$

$\Rightarrow$ Robustness against outliers