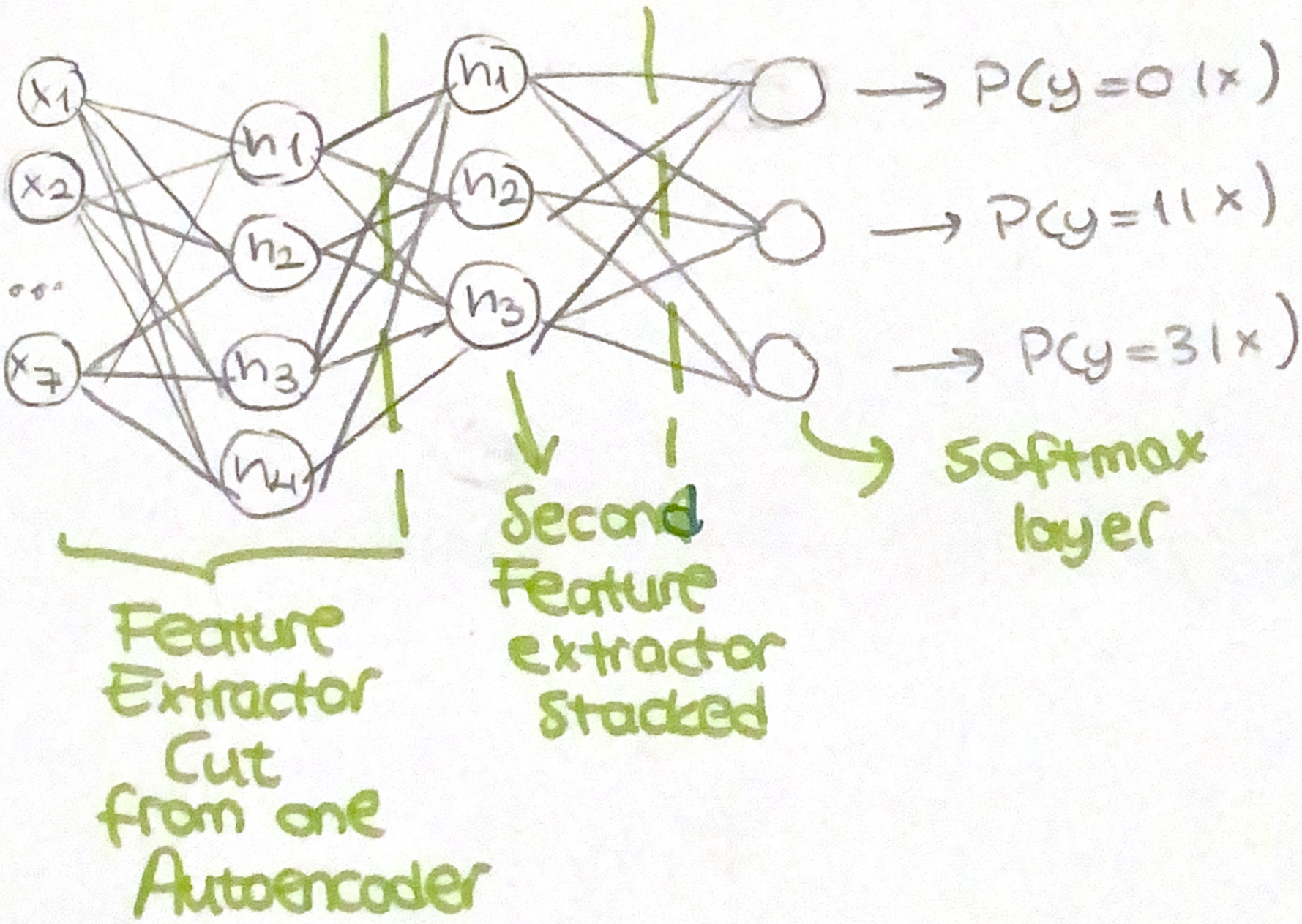
Autoencoders (contd.)

Stacked Autoencoder

1. Train autoencoders to be feature extractor,
2. Add couple of more layers (optional)
3. Add a classifier layer on top → Learning becomes supervised
4. Train with domain specific data

→ This is used to avoid vanishing gradients.



Denoising Autoencoder

Changing the reconstruction error term.

Traditional reconstruction loss → $L(x, \bar{x})$ ($\bar{x}$ = decoder output)

Denoiser autoencoder loss → $L(x, g(f(\tilde{x}))$

$g(f(x))$ = hidden layer output

$\tilde{x}$ = corrupted input

It receives a corrupted datapoint as input and tries to predict uncorrupted original data point.

$C(\tilde{x}|x)$ → conditional distr over corrupted samples ($\tilde{x}$)

Preconstruct $(x|\tilde{x})$ → reconstruction distribution estimated from training pairs.

$\text{Preconstr}(x|\tilde{x}) = \underbrace{p_{\text{decoder}}(x|h)}_{g(h)}$ → output of encoder $f(\tilde{x})$