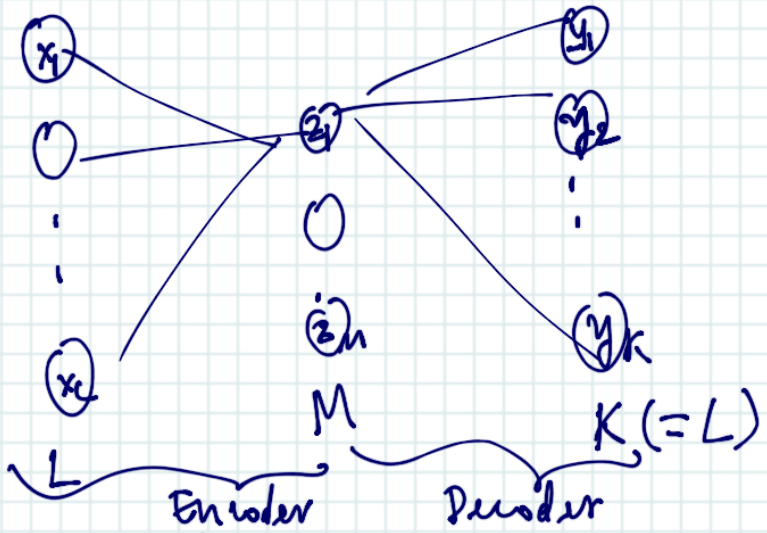
14/10/19

EE5301: Representation Learning

- Review
- Autoencoders (AE)
- AE with sparsity
- **Autoencoder (AE):**
  - Learn hidden/efficient representation of input (training set)
  - Uses a standard MLP framework with labels being the input



  - Constraint: $M < L$.

- **Sparse AE:** Recall: $z_m^{(i)} = \sigma\left(\underline{\alpha}_m^T \cdot \underline{x}^{(i)} + \alpha_{m0}\right)$,

  Also, $\bar{z}_m = \frac{1}{N} \sum_{i=1}^{N} z_m^{(i)}$. We would like z_m to be a Bernoulli RV with prob. of $1 = \rho$.

  Impose this sparsity constraint using the Kullback-Leibler divergence between Bern(ρ) and z_m.

$$KL(p \| q) = \sum_{x \in X} p(x) \cdot \log\left[\frac{p(x)}{q(x)}\right]$$

For our case, $KL(Bern(p) \| \bar{z}_m) = p \cdot \log \frac{p}{\bar{z}_m} +$

$$(1-p) \cdot \log \frac{(1-p)}{(1-\bar{z}_m)} \quad \text{--- (1)}$$

We will use this term to "regularize" our network, i.e.,

$$R_s(\theta) = R(\theta) + \lambda \cdot \sum_{m=1}^{M} KL(Bern(p) \| \bar{z}_m)$$