Entropy and Information.

$$H[p(x)] = E_{p(x)}[-\log p(x)]$$

$$X \in \{x_1, x_2, x_3, \ldots\}$$

$$H[p(x)] = \sum_{x_i \in X} -p(x_i) \log p(x_i)$$

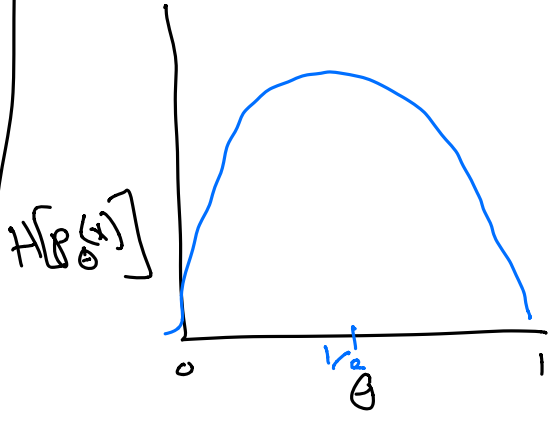
Bernoulli:

$$x \in \{0, 1\}$$

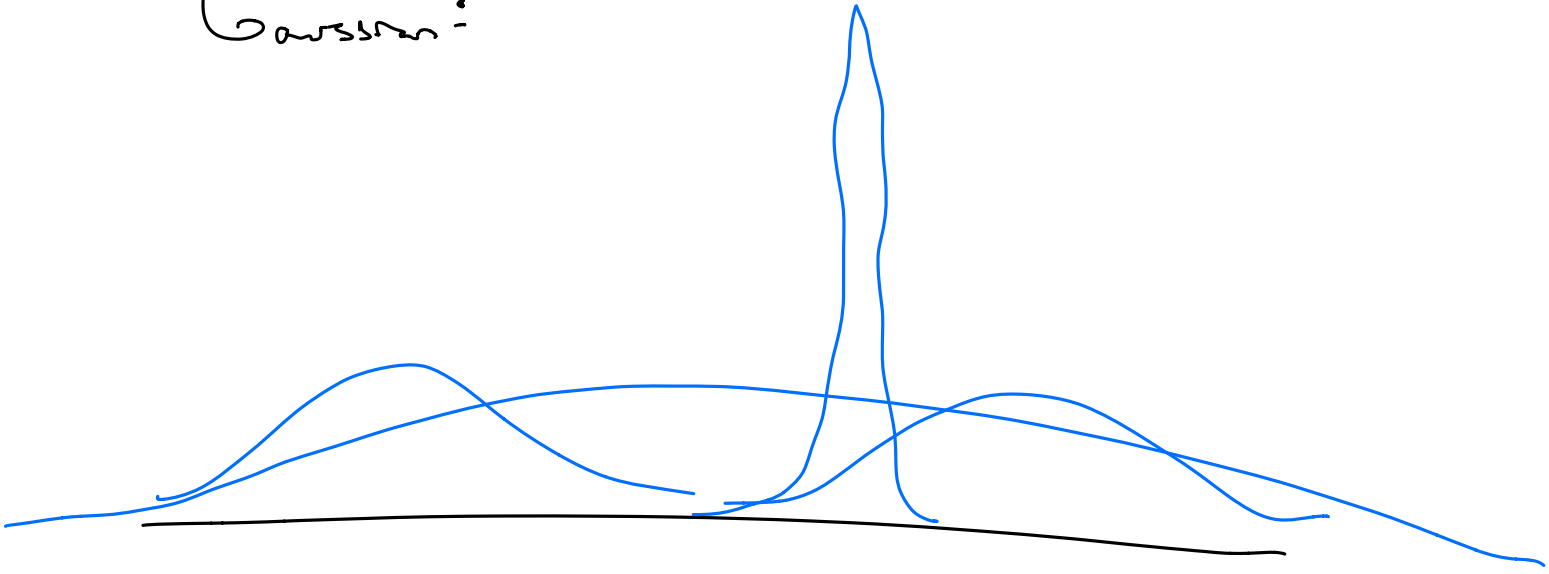
$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$

$$\theta \in [0, 1]$$

$$H = -\theta \log \theta - (1-\theta)\log(1-\theta)$$



Gaussian:



Mutual Information

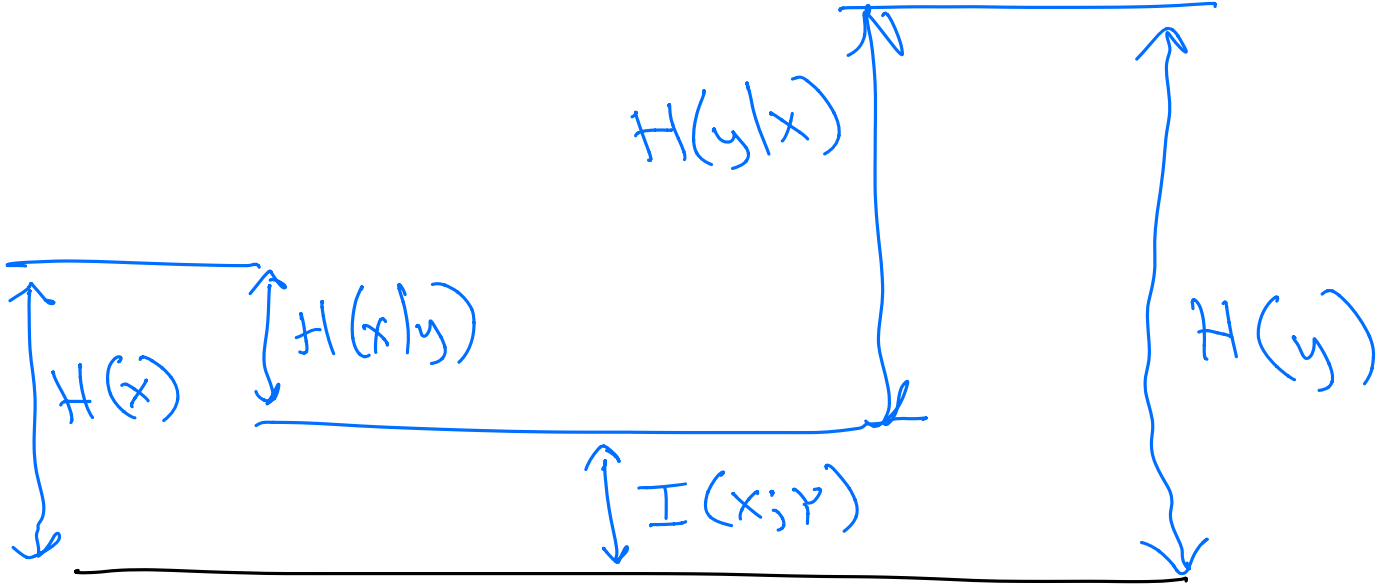
r.v. X, Y

$$I(X;Y) = D_{KL}\left(p(x,y) \,\|\, p(x)p(y)\right)$$

$$= E_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] = E\left[\log \frac{p(x|y)}{p(x)}\right] = E\left[\log \frac{p(y|x)}{p(y)}\right]$$

$$= H[X] - H[X|Y]$$

$$= H[Y] - H[Y|X]$$



X

sender

Y = noisy(X)

receiver

Active Learning

Labels: $\{x_i, y_i\}_{i=1}^{N} \equiv \mathcal{D}$ $\quad (M >> N)$

Unlabelled ("pool") $\{x_j\}_{j=1}^{M} \equiv \mathcal{U}$

Q: which $x_i \in \mathcal{U}$ should we label next?

$y \in \{1, \ldots, K\}$

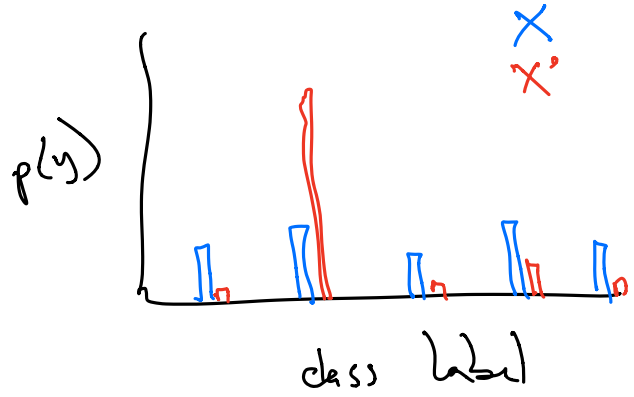
ML: $x \rightarrow p(y|x) = \text{Categorical}[1, \ldots, K]$

$= [\alpha_1, \ldots, \alpha_K]$

$p(y=1|x)$

Consider candidate $x^* \in \mathcal{U}$.

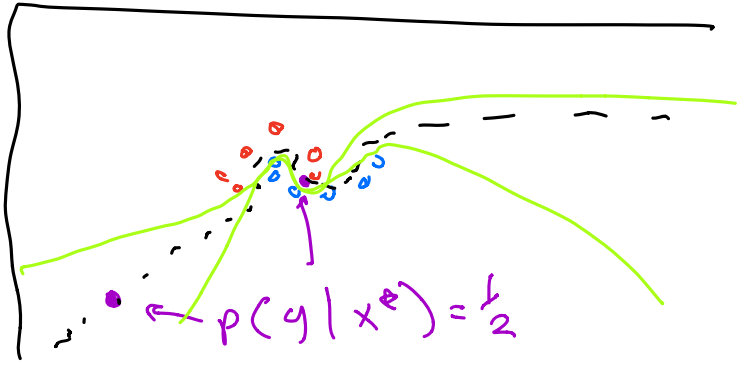
$$\max_{\alpha_k} \underbrace{p(y \mid x^*)}_{J(x^*)} \; ?$$



Algo:

$$\operatorname*{argmin}_{x^* \in \mathcal{U}} \; \overline{J}(x^*)$$

$$\operatorname*{argmax}_{x^* \in \mathcal{U}} \; H(p(y \mid x^*))$$



Bayesian A.C. by Disagreement. (BALD)

$$p(y|x^*, \mathcal{D}) = \int p(y|x^*, \theta) p(\theta|\mathcal{D}) d\theta$$

$$\operatorname{argmax}_{x^* \in \mathcal{U}} I(y, \theta | \mathcal{D}, x^*) = E_{p(\theta, y|\mathcal{D}, x^*)} \left[\log \frac{p(\theta|y, x^*, \mathcal{D})}{p(\theta|\mathcal{D})} \right]$$

$$= H[\theta|\mathcal{D}] - E_{p(y|\mathcal{D}, x^*)} [H[\theta|y, x^*, \mathcal{D}]]$$

$$= H[y|x^*, \mathcal{D}] - E_{p(\theta|\mathcal{D})} [H[y|x^*, \theta]]$$

$$H[y|x^*, \mathcal{D}] = H[E_{p(\theta|\mathcal{D})}[p(y|x^*, \theta)]]$$

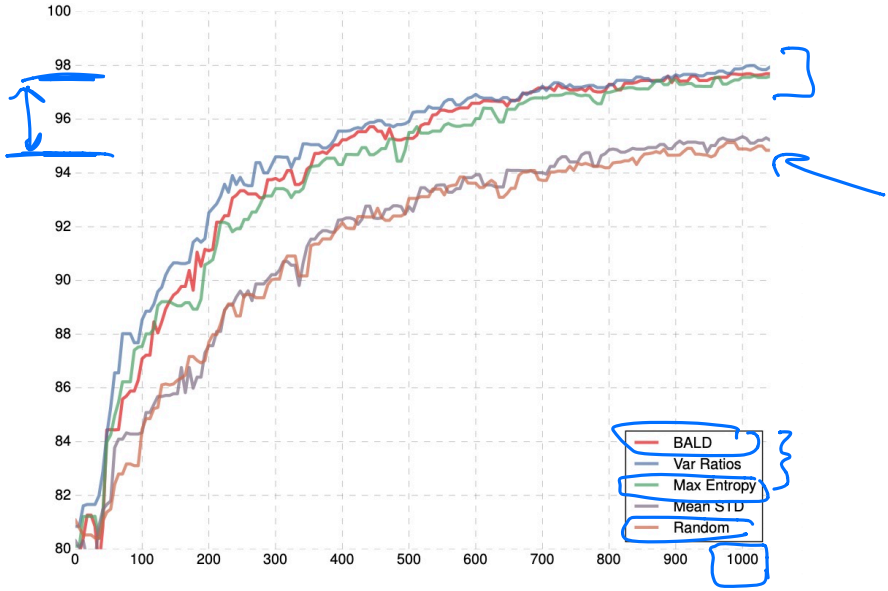


Figure 1. **MNIST test accuracy as a function of number of acquired images from the pool set** (up to 1000 images, using validation set size 100, and averaged over 3 repetitions). Four acquisition functions (*BALD*, *Variation Ratios*, *Max Entropy*, and *Mean STD*) are evaluated and compared to a *Random* acquisition function.

[Gal et al., 2017]