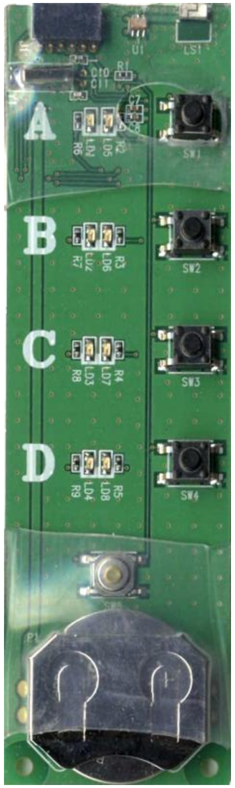
# EE488 Special Topics in EE <Deep Learning and AlphaGo>

Sae-Young Chung

Lecture 3 Supplementary Material

September 6, 2017

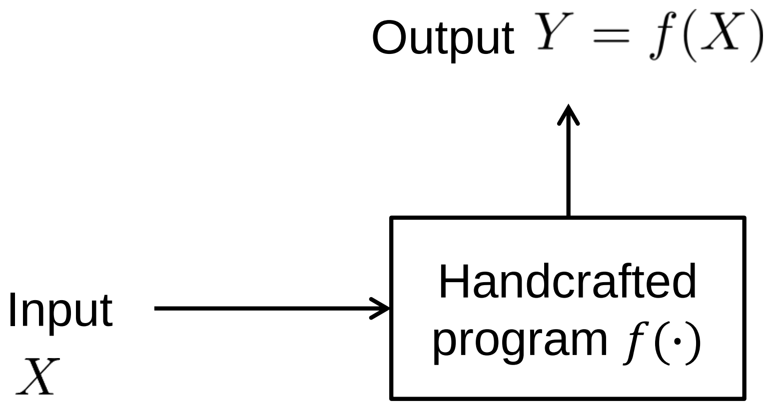
- Clicker
  - Press one of four buttons (A, B, C, or D)
  - Your answer is locked in once pressed, so be careful when pressing a button
  - Initially a red light next to the pressed button will start flashing
  - It will turn green if your response is confirmed
  - All lights will be automatically turned off once data collection is finished, therefore you don’t need to do anything while red or green lights are flashing

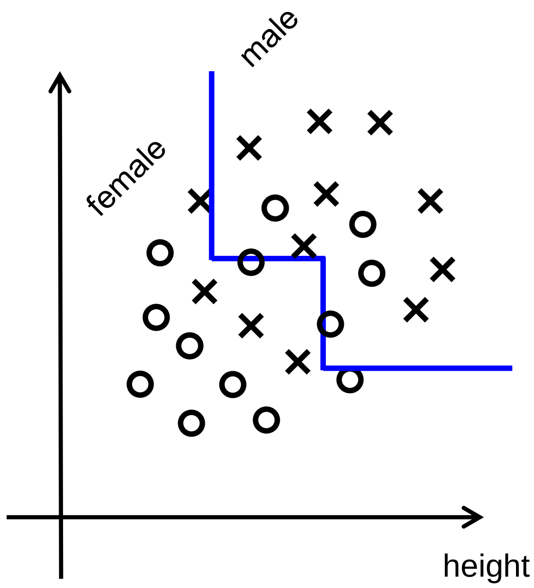


# Rule-based System Example

Example) X=('height', 'weight')

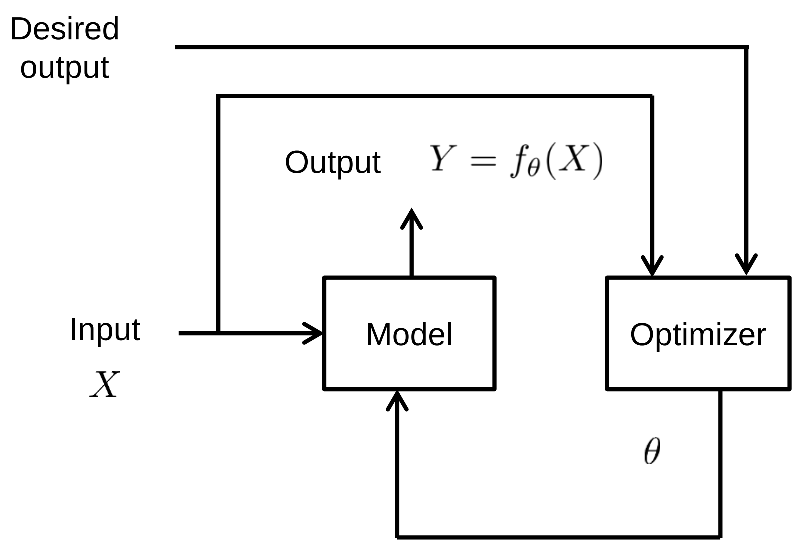
Y='male' if 'height' > 160cm & 'weight' > 70kg
or 'height' > 170cm & 'weight' > 60kg
Y='female' otherwise

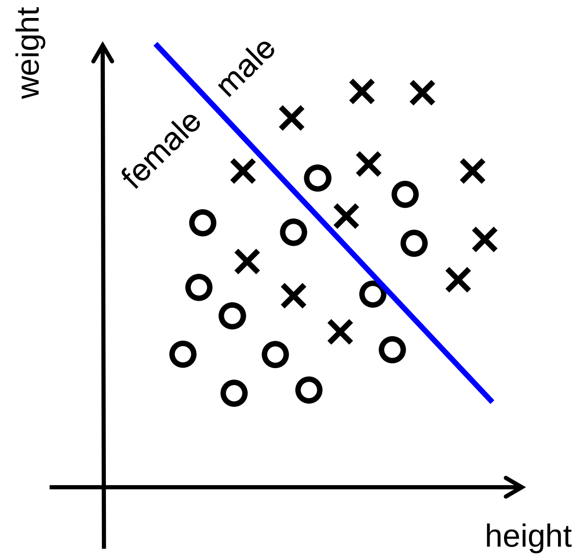


# Classical Machine Learning Example

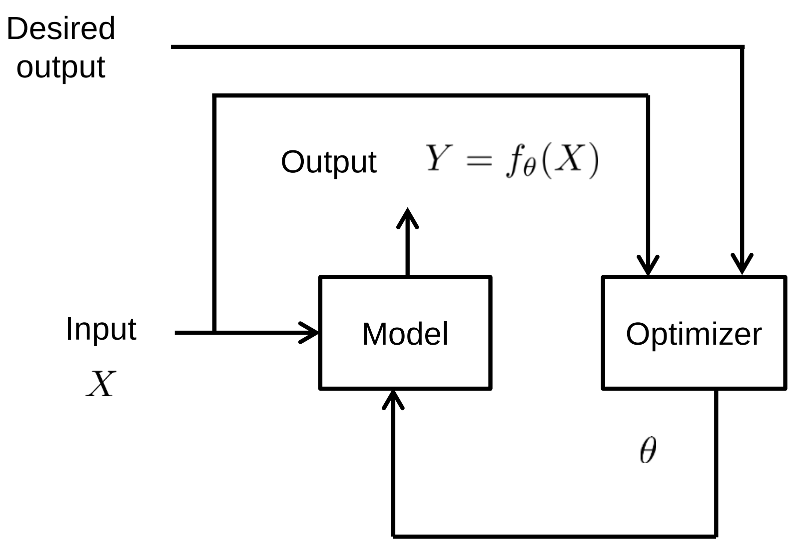
- $Y$ = 'male' if $\theta_0 *$'height'$+\theta_1 *$'weight'$> 0$ and $Y$ = 'female' otherwise
- $\theta = (\theta_0, \theta_1)$ is learned from training data to minimize error
- Features are 'height' and 'weight' and they are NOT learned from data
- How to learn? Support vector machine (SVM), ...

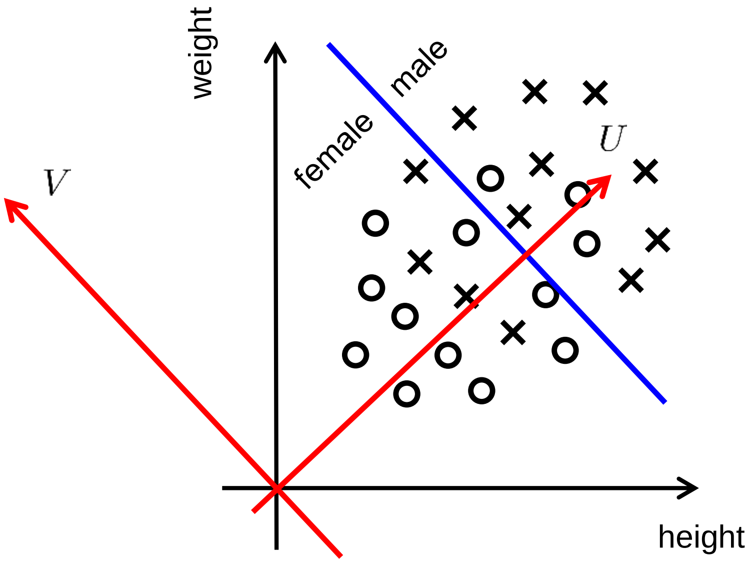


# Representation Learning Example

- $Y$ = 'male' if $U > \theta$ and $Y$ = 'female' otherwise
- $\theta$ is learned from training data to minimize error
- Representation from 'height' and 'weight' to $(U, V)$ is also learned from training data, e.g., principal component analysis (PCA), ...



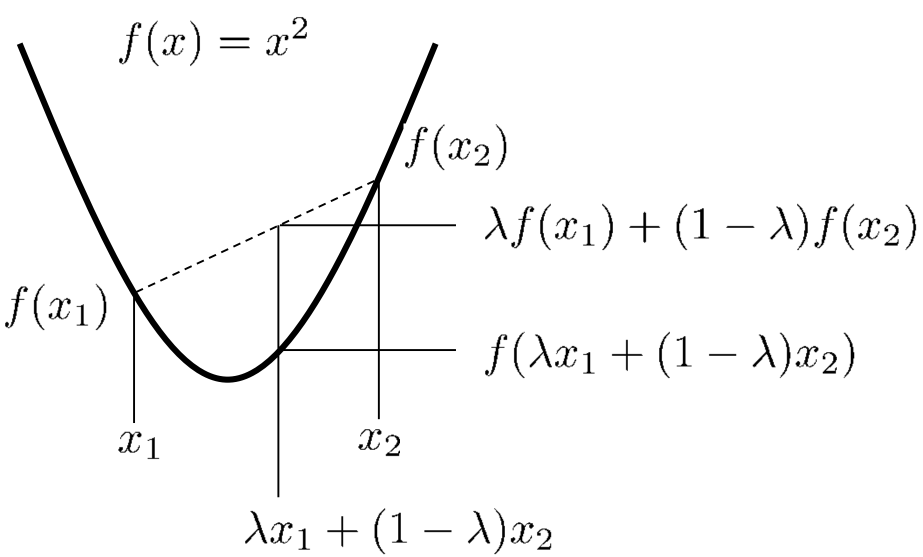



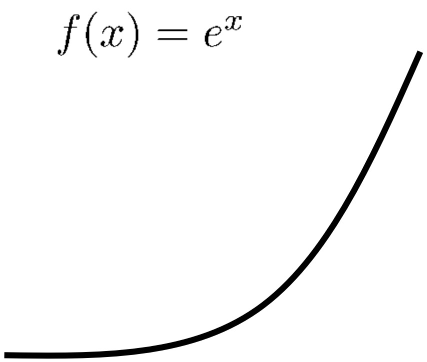
# Convex Function

- A function $f(x)$ is *convex* over an interval $(a, b)$ if for every $x_1, x_2 \in (a, b)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

- A function $f(x)$ is *strictly convex* if equality holds only if $\lambda = 0, 1$.
- A function $f(x)$ is *concave* if $-f(x)$ is convex.
- Examples of convex functions.



# Jensen's Inequality

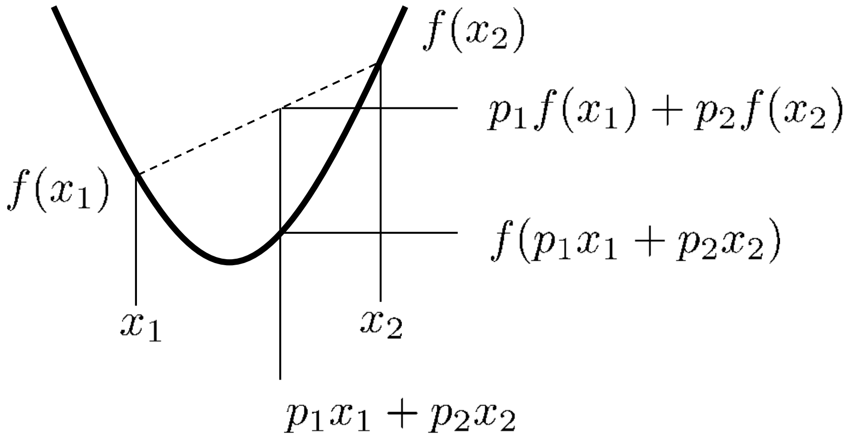
- *Thm.* (Jensen's inequality) If $f$ is a convex function and $X$ is a random variable, then

$$E[f(x)] \geq f(E[X])$$

If $f$ is strictly convex, the equality holds iff $X = E[X]$ with probability 1.
*Proof* If $X$ is two-valued, i.e., $x_1$ and $x_2$ with probabilities $p_1$ and $p_2$, respectively, then from convexity we get

$$p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2).$$

Can be generalized to multi-valued distributions and to continuous distributions.



# Non-Negativity of Relative Entropy

- *Thm.* If $p(x)$ and $q(x)$ are two pmf's, then

$$D(p\|q) \geq 0$$

  with equality iff $p(x) = q(x)$ for all $x$.

- *Proof)* Let $A = \{x : p(x) > 0\}$.

$$\begin{aligned} -D(p\|q) &= \sum_{x \in A} p(x) \log \frac{q(x)}{p(x)} \\ &\leq \log \sum_{x \in A} p(x) \frac{q(x)}{p(x)} \\ &= \log \sum_{x \in A} q(x) \\ &\leq \log \sum_{x \in \mathcal{X}} q(x) \\ &= \log 1 = 0 \end{aligned}$$

# Non-Negativity of Mutual Information

- *Corollary* (Non-negativity of mutual information) For any two random variables $X$ and $Y$,
$$I(X;Y) \geq 0,$$
with equality iff $X$ and $Y$ are independent.
*Proof* $I(X;Y) = D(p(x,y)\|p(x)p(y)) \geq 0$ with equality iff $p(x,y) = p(x)p(y)$.

- *Corollary* (Conditioning reduces entropy) $H(X|Y) \leq H(X)$ with equality iff $X$ and $Y$ are independent.

- *Corollary*
$$D(p(y|x)\|q(y|x)) \geq 0,$$
with equality iff $p(y|x) = q(y|x)$ for all $x$ and $y$ such that $p(x) > 0$.

- *Corollary*
$$I(X;Y|Z) \geq 0,$$
with equality iff $X$ and $Y$ are conditionally independent given $Z$.