Lecture 11.

Sufficient Statistic

Y : observation, infer about parameter X
$\uparrow$ high-dimensional

find $f(Y)$ - low-dimensional, use $f(Y)$ to infer/predict X.

$$Y \rightarrow \boxed{NN} \rightarrow f(Y) \rightarrow X$$

Dimension reduction / Feature selection

Non Bayesian case :

Observation Y, parameter X, model $P_Y(y;x)$

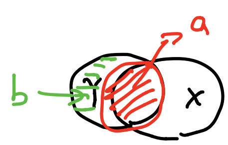
statistic of Y, $t(Y)$ is a function of Y, could be a many to one function

__Definition__ : $t(y)$ is a sufficient statistic (s.s.) for a model $P_Y(y;x)$ if $P_{Y|T}(y|t;x)$ is NOT a function of x.

Note that $P_Y(y;x) = \sum_t \underset{\substack{\uparrow \\ P_{YT}(y,t(y');x)=0 \\ \text{if } y'\neq y}}{P_{YT}(y,t;x)} \underset{\substack{\uparrow \\ t=t(y)}}{=} P_{YT}(y,t(y);x) = P_{Y|T}(y|t(y);x)\cdot P_T(t(y);x)$

$$\underset{\substack{\uparrow \\ \text{if } t(y) \text{ is a s.s.}}}{=} \underbrace{P_{Y|T}(y|t(y))}_{\text{func. of } Y} \cdot \underbrace{P_T(t(y);x)}_{\text{func. } t(y),\, x}$$

__Thm__ : $t(y)$ is a s.s. if and only if $\dfrac{P_Y(y;x)}{P_T(t(y);x)}$ is not a function of x, $\forall x, y$.



__Thm__ : (Fisher - Neyman Factorization Theorem)

A statistic $t(y)$ is a s.s. iff $P_Y(y;x) = a(t(y),x)\cdot b(y)$, $\forall x,y$)

Proof: "$\Rightarrow$" If $t(y)$ is a s.s., then take $a(t(y), x) = P_T(t(y); x)$, $b(y) = P_{Y|T}(y|t(y))$

"$\Leftarrow$" If $P_Y(y;x) = a(t(y), x) \cdot b(y)$

then $$P_T(t;x) = \sum_{y:\,t(y)=t} P_Y(y;x) = \Big[\sum_{y:\,t(y)=t} b(y)\Big] \cdot a(t, x)$$

$$P_{Y|T}(y|t;x) = \frac{P_{YT}(y,t;x)}{P_T(t;x)} = \frac{P_Y(y;x)\cdot \overset{1}{\cancel{P_{T|Y}(t|y;x)}}}{P_T(t;x)} \underset{\substack{\uparrow \\ t=t(y)}}{=} \frac{P_Y(y;x)}{P_T(t;x)} = \frac{\cancel{a(t,x)}\cdot b(y)}{\cancel{a(t,x)}\cdot \sum_{y:\,t(y)=t} b(y)}$$

$$= \frac{b(y)}{\sum_{y:\,t(y)=t} b(y)} \quad \text{is NOT a function of } x \qquad \square$$

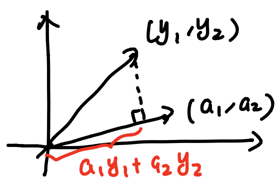
Example: $Y_1 \sim N(a_1 x, 1)$, $Y_2 \sim N(a_2 x, 1)$, Y_1, Y_2 are indep.

$$P_{Y_1Y_2}(y_1, y_2; x) = \frac{1}{2\pi} \exp\Big[-\frac{1}{2}(y_1 - a_1 x)^2 - \frac{1}{2}(y_2 - a_2 x)^2\Big]$$

$$= \underbrace{\frac{1}{2\pi}\exp\Big[-\frac{1}{2}(y_1^2 + y_2^2)\Big]}_{b(y_1, y_2)} \cdot \underbrace{\exp\Big[x\cdot(a_1 y_1 + a_2 y_2) - \frac{1}{2}(a_1^2 + a_2^2)x^2\Big]}_{a(t,\,x)}$$

$T = \underline{a_1 Y_1 + a_2 Y_2}$ is a s.s.

$t(y_1, y_2) = a_1 y_1 + a_2 y_2$ is a s.s.



Example: $t(y) = y$ is a s.s.

Example: $X \in \{H_0, H_1\}$ binary hypothesis testing problem.

$$t(y) = \frac{P_Y(y;H_1)}{P_Y(y;H_0)} \quad \text{is a s.s.}$$

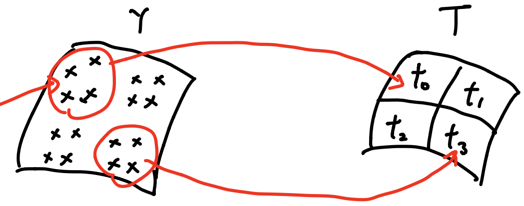
since $$\underset{\substack{\uparrow \\ H = H_0 \text{ or } H_1}}{\underline{P_Y(y;H)}} = \underbrace{\big[P_Y(y;H_0) + P_Y(y;H_1)\big]}_{b(y)} \cdot a(t, H) \qquad a(t;H) = \begin{cases} \frac{1}{1+t} & \text{if } H = H_0 \\ \frac{t}{1+t} & \text{if } H = H_1 \end{cases}$$

Example: Exponential family $P_Y(y;x) = \exp[\lambda(x)\cdot t(y) - \alpha(x) + \beta(y)]$

$a(t, x) = \exp[\lambda(x)\, t(y) - \alpha(x)]$, $b(y) = \exp[\beta(y)]$

$t(y)$ is a s.s.
$\uparrow$
natural statistic

Partition : $Y \to t(Y)$ is a many-to-one function



partition of the values Y can take $\{y : t(y) = t_0\}$

if y_1, y_2 s.t. $t(y_1) = t(y_2)$, then y_1, y_2 are in the same partition.

<u>Thm</u>: $t(y)$ is a s.s. iff for all y_1, y_2 in the same partition (i.e, $t(y_1) = t(y_2)$)

$\exists\ g(y_1, y_2)$ s.t. $P_Y(y_1; x) = g(y_1, y_2) \cdot P_Y(y_2; x)$, $\forall x$

That is, $\dfrac{P_Y(y_1; x)}{P_Y(y_2; x)}$ is NOT a function of x

<u>Proof</u>: "$\Rightarrow$" if t is a s.s. $\dfrac{P_Y(y_1; x)}{P_Y(y_2; x)} = \dfrac{a(\cancel{t(y_1)}, x) \cdot b(y_1)}{a(\cancel{t(y_2)}, x) \cdot b(y_2)} = \dfrac{b(y_1)}{b(y_2)} = g(y_1, y_2)$

"$\Leftarrow$" for any t_0 and all y s.t. $t(y) = t_0$, we have

$$P_{Y|T}(y | t_0; x) = \frac{P_{YT}(y, t_0; x)}{P_T(t_0; x)} = \frac{P_Y(y; x)}{\sum\limits_{y' : t(y') = t_0} P_Y(y'; x)} = \frac{1}{\sum\limits_{y' : t(y') = t_0} \underbrace{\frac{P_Y(y'; x)}{P_Y(y; x)}}_{g(y', y)}} = \frac{1}{\sum\limits_{y' : t(y') = t_0} g(y', y)}$$

NOT function of x

$\square$

Minimal Sufficient Statistic

<u>Definition</u>: A s.s. $t^*(y)$ is minimal if for any other s.s. $t(y)$, $t^* = g(t)$ for some function $g(\cdot)$

<u>Remarks</u>: (1) t^* is not unique, any 1-to-1 function g, $g(t^*)$ is minimal

(2) All information in t^* are useful, cannot further reduce information.

<u>Definition</u>: A s.s. t is complete if for any function $\phi(\cdot)$

$$\mathbb{E}_x[\phi(t(Y))] = 0, \forall x \Rightarrow \phi(\cdot) = 0$$

$$\sum_y P_Y(y; x) \cdot \phi(t(y))$$

<u>Thm</u>: If a s.s. t is complete, then t is minimal.

proof: suppose $S(y)$ is a minimal s.s., then exists $g(\cdot)$ s.t. $S = g(t)$

Note that $\mathbb{E}[T] = \mathbb{E}[\underbrace{\mathbb{E}[T|S]}_{\text{function of } S}]$, note that s is a func. of $t \Rightarrow \mathbb{E}[T|S]$ is a func. of T.

Let's define $\phi(T) = T - \mathbb{E}[T|S] \Rightarrow \mathbb{E}_x[\phi(T)] = \mathbb{E}[T] - \mathbb{E}[\cancel{\mathbb{E}[T|S]}^{\mathbb{E}[T]}] = 0, \ \forall x$

$\Rightarrow \phi(\cdot) = 0 \Rightarrow T = \mathbb{E}[T|S] \Rightarrow T$ is a func. of S

$\Rightarrow T$ is minimal $\square$

Example: binary hypothesis testing, $\mathcal{X} = \{H_0, H_1\}$, s.s. $t(y) = \dfrac{P_Y(y, H_1)}{P_Y(y, H_0)}$

Then, $t(y)$ is a minimal s.s.

show that $t(y)$ is complete (Homework).

Bayesian case.

Observe Y, and X is a random variable, model $P_{XY}(x,y) = P_{Y|X}(y|x) \cdot P_X(x)$

Definition: $t(y)$ is a s.s. w.r.t. P_{XY} if $P_{Y|TX}(y|t(y), x) = P_{Y|T}(y|t(y))$

Thm: $t(y)$ is a s.s. iff $P_{X|Y}(x|y) = P_{X|T}(x|t(y))$, $\forall x, y$.

Proof: "$\Rightarrow$" if t is a s.s., then

$$P_{X|Y}(x|y) = P_{X|YT}(x|y, t(y)) = \frac{P_{XYT}(x, y, t(y))}{P_{YT}(y, t(y))} = \frac{\cancel{P_T(t(y))} \cdot P_{X|T}(x|t(y)) \cdot \cancel{P_{Y|XT}(y|x, t(y))}}{\cancel{P_T(t(y))} \cdot \cancel{P_{Y|T}(y|t(y))}}$$

$$= P_{X|T}(x|t(y))$$

"$\Leftarrow$" if $P_{X|Y}(x|y) = P_{X|T}(x|t(y))$, then

$$P_{Y|TX}(y|t(y), x) = \frac{P_{YTX}(y, t(y), x)}{P_{TX}(t(y), x)} = \frac{\cancel{P_T(t(y))} \cdot P_{Y|T}(y|t(y)) \cdot \overset{\cancel{P_{X|Y}(x|y)}}{P_{X|YT}(x|y, t(y))}}{\cancel{P_T(t(y))} \cdot \underset{\cancel{P_{X|Y}(x|y)}}{P_{X|T}(x|t(y))}} = P_{Y|T}(y|t(y))$$

$\square$

Thm: (Factorization): $t(y)$ is a s.s. iff $P_{Y|X}(y|x) = \underbrace{P_{T|X}(t(y)|x)}_{a(t,x)} \cdot \underbrace{P_{Y|T}(y|t(y))}_{b(y)}$

proof: Note that $P_{Y|X}(y|x) = P_{YT|X}(y, \underset{\uparrow \text{ func. of } y.}{t(y)}|x) = P_{T|X}(t(y)|x) \cdot \underbrace{P_{Y|TX}(y|t(y), x)}_{= P_{Y|T}(y|t(y))}$

since t is a s.s $\square$

Definition: $X \longleftrightarrow Y \longleftrightarrow Z$ is a Markov chain if X, Z are conditionally indep. given Y

$\Rightarrow \quad \underset{= P_{X|Y} \cdot P_{Z|XY}}{P_{XZ|Y}} = P_{X|Y} \cdot P_{Z|Y}$

$\Rightarrow P_{Z|XY} = P_{Z|Y}$

$$P_{XYZ} = P_X \cdot P_{Y|X} \cdot P_{Z|XY} \underset{\uparrow}{=} P_X \cdot P_{Y|X} \cdot P_{Z|Y}$$

if $X \leftrightarrow Y \leftrightarrow Z$ M.C.

Then, $t(y)$ is a s.s. iff $X \leftrightarrow T \leftrightarrow Y$ is a Markov chain.

Note that $X \leftrightarrow Y \leftrightarrow T$ is always a M.C.

Data processing inequality: $I(X;Y) \geq I(X;T)$

If T is a s.s. then "=" holds.