# Hands on session 4: Introduction to supervised learning

## Conditional Expectation:

Supervised learning setup:

$$(X, Y) \sim \mathcal{L}$$

input: $((X_i, Y_i))_i$ iid from $\mathcal{L}$

Can I predict $Y$ from $X$ using the examples?

In other words, can I estimate $P(Y \mid X)$?

Can I have access to the moments of $Y$ when I have access to $X$!

$$\boxed{E_{\mathcal{L}}(Y \mid X=x) = E_{p(\cdot \mid X=x)}(Y)}$$

**Theorem - Definition:** Let $\mathcal{B} \subset \mathcal{A}$ be a $\sigma$-algebra, let $X \in L^1(\Omega, \mathcal{A}, P)$. There exists a unique random variable in $L^1(\Omega, \mathcal{B}, P)$ noted $E(X \mid \mathcal{B})$ st

$$\forall B \in \mathcal{B}, \; E(X \mathbb{1}_B) = E(E(X \mid \mathcal{B}) \mathbb{1}_B)$$

Furthermore, we have for every $Z$ $\mathcal{B}$-measurable bounded,

$$E(XZ) = E(E(X \mid \mathcal{B}) Z)$$

If $X \geq 0$, $E(X \mid \mathcal{B}) \geq 0$

**Definition:** Let $X, Y$ be two r.v in $L^1(\Omega, \mathcal{A}, P)$

$$E(X \mid Y) \doteq E(X \mid \sigma(Y))$$

**Properties:** (a) If $X$ is $\mathcal{B}$ measurable, $E(X|\mathcal{B}) = X$
(b) $X \mapsto E(X|\mathcal{B})$ is linear
(c) $E(E(X|\mathcal{B})) = E(X)$
(d) $|E(E(X|\mathcal{B}))| \leq E(|X||\mathcal{B})$ a.s. So, $E(|E(X|\mathcal{B})|) \leq E(X)$
(e) $X \geq X' \Rightarrow E(X|\mathcal{B}) \geq E(X'|\mathcal{B})$ a.s.

**Proposition:** $X$ rv, $Y$ rv $\mathcal{B}$ measurable $E(YX|\mathcal{B}) = Y E(X|\mathcal{B})$ if well defined.

**Proposition:** $\mathcal{B}_1, \mathcal{B}_2$ $\sigma$-algebras st $\mathcal{B}_1 \subset \mathcal{B}_2$
then, $\forall X \in L^1$, $E(E(X|\mathcal{B}_2)|\mathcal{B}_1) = E(X|\mathcal{B}_1)$

**Theorem** $\mathcal{B}_1, \mathcal{B}_2$ $\sigma$-algebras,

$$\mathcal{B}_1 \perp \mathcal{B}_2 \iff \forall X \in L^1 \ \mathcal{B}_2 \text{ measurable}, \quad E(X|\mathcal{B}_1) = E(X)$$

**Theorem:** $X \in L^2(\Omega, \mathcal{A}, P)$, then $E(X|\mathcal{B}) = \text{proj}^{\perp}_{L^2(\Omega,\mathcal{B},P)}(X)$

which means that
$$\boxed{E(X|\mathcal{B}) = \underset{X' \mathcal{B} \text{ measurable}}{\operatorname{argmin}} E\left((X - X')^2\right)}$$

**Proposition:** For every borelian function $h$,

$$E(h(Y)|X=x) = \begin{cases} \frac{1}{q(x)} \int h(y) p(x,y)\, dy & \text{if } q(x) \neq 0 \\ \text{whatever} & \text{otherwise} \end{cases}$$

where $q(x) = \int p(x,y)\, dy$

For proofs, see "Intégration, Probabilités et Processus Aléatoires" from Jean-François Le Gall.

## Exercise 5. Classification 1:

$$\ell(\hat{y}, y) = \begin{cases} c \text{ if } \hat{y} = 1,\ y = 0 \\ 1 \text{ if } \hat{y} = 0,\ y = 1 \\ 0 \text{ otherwise} \end{cases}$$

1. $L(g) = E(\ell(g(X), Y)) = cP(g(X) = 1, Y = 0) + P(g(X) = 0, Y = 1)$

2. $g^* = \underset{g}{\operatorname{argmin}}\, L(g) = E(\ell(g(X), Y))$

$$E(\ell(g(X), Y)) = E_X \underbrace{E(\ell(g(X), Y) \mid X)}_{\text{minimize}}$$

$$E(\ell(g(X), Y) \mid X) = E\big(c\, g(X)(1 - Y) + (1 - g(X)) Y \mid X\big)$$

$$= c g(X)(1 - E(Y \mid X)) + (1 - g(X)) E(Y \mid X)$$

$$= g(X)\big[c - (1 + c) E(Y \mid X)\big] + E(Y \mid X)$$

So we want $g(X) = 1$ when $\frac{c}{1+c} \leq E(Y \mid X)$

So $f^*(x) = \mathbb{1}_{\eta(x) \geq \frac{c}{1+c}}$ where $\eta(x) = E(Y \mid X = x)$
$= P(y = 1 \mid x = x)$

---

## Exercise 6: Classification 2

1. $|L(g) - L(g')| = |E(\ell(g(X), Y)) - E(\ell(g'(X), Y))|$

$$\leq E\big(\underbrace{|\ell(g(X), Y) - \ell(g'(X), Y)|}_{= \mathbb{1}_{g(X) \neq g(X')}}\big)$$

So, $|L(g)-L(g')| \leq E(\mathbb{1}_{g(x)\neq g'(x)}) = \mathbb{P}(g(x)\neq g'(x))$

2. $L(g) = E(\ell(g(x), Y)) = E_x\, E(\ell(g(x), Y) \mid X)$

$= E_X\left(g(x)\left(1-2E(Y\mid X)\right) + E(Y\mid X)\right)$

$= E_X\left((1-\mathbb{1}_{g(x)\neq 1})(1-2E(Y\mid X)) + E(Y\mid X)\right)$

$= E_X\left(\mathbb{1}_{g(x)\neq 1}(2E(Y\mid X)-1) + (1-E(Y\mid X))\right)$

So, $L(g) = E\left(\mathbb{1}_{g(x)\neq 1}(2\eta(X)-1) + (1-\eta(X))\right)$

3. $|L(g)-L(g')| = \left|E\left((\mathbb{1}_{g(x)\neq 1} - \mathbb{1}_{g'(x)\neq 1})(2\eta(X)-1)\right)\right|$

$\leq E\left(\underbrace{|\mathbb{1}_{g(x)\neq 1} - \mathbb{1}_{g'(x)\neq 1}|}_{=\mathbb{1}_{g(x)\neq g'(x)}}\, |2\eta(X)-1|\right)$

So $|L(g)-L(g')| \leq E\left(|2\eta(X)-1|\,\mathbb{1}_{g(X)\neq g'(X)}\right)$

4. $x \in G \Delta G^*$ iff $g(x)\neq g^*(x)$

So $|L(g)-L(g^*)| \leq E\left(|2\eta(X)-1|\,\mathbb{1}_{G\Delta G^*}(X)\right)$

Furthermore, by definition, $L(g) \geq L(g^*)$

So $L(g)-L(g^*) \leq E\left(|2\eta(X)-1|\,\mathbb{1}_{G\Delta G^*}(X)\right)$