Remarks on Notations

MATH-412 - Statistical Machine Learning

Notations

x, y	plain lower case letter denote scalars or elements from general spaces
$\mathbf{x}$	bold lower case denote vectors (fixed or random depending on context)
X	capitals denote a scalar or vectorial random variable
$\{X = \mathbf{x}\}$	denotes the event that the random variable X takes the value $\mathbf{x}$
$\boldsymbol{X}$	bold capitals denote matrices (possibly random)
D_n	denotes the training set $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$.
$\widehat{f}, \widetilde{f}$	random predictor which depend on D_n.

Remarks on conditional expectations

$$\mathcal{R}(f) = \int \ell(f(\mathbf{x}), y)\, dP(\mathbf{x}, y) = \mathbb{E}[\ell(f(X), Y)]$$

$$\mathcal{R}(\widehat{f}) = \int \ell(\widehat{f}(\mathbf{x}), y)\, dP(\mathbf{x}, y) = \mathbb{E}[\ell(f(X), Y)|D_n]$$

$$\mathbb{E}[\mathcal{R}(\widehat{f})] = \int \ell(\widehat{f}(\mathbf{x}), y)\, dP(\mathbf{x}, y) \prod_{i=1}^{n} dP(\mathbf{x}_i, y_i) = \mathbb{E}[\ell(\widehat{f}(X), Y)]$$

$$\begin{aligned}
\mathcal{R}(a|\mathbf{x}) &= \mathbb{E}[\ell(\quad a \quad, Y)|X = \mathbf{x}] \\
\mathcal{R}(f(X)|X) &= \mathbb{E}[\ell(f(X), Y)|X] \\
\mathcal{R}(f(\mathbf{x})|\mathbf{x}) &= \mathbb{E}[\ell(f(\mathbf{x}), Y)|X = \mathbf{x}] \\
\mathcal{R}(\widehat{f}(\mathbf{x})|\mathbf{x}) &= \mathbb{E}[\ell(\widehat{f}(\mathbf{x}), Y)|X = \mathbf{x}, D_n]
\end{aligned}$$

Alternate notations with partial expectations

$$\mathcal{R}(f) = \int \ell(f(\mathbf{x}), y)\, dP(\mathbf{x}, y) = \mathbb{E}[\ell(f(X), Y)] \qquad = \mathbb{E}_{X,Y}\left[\ell(f(X), Y)\right]$$

$$\mathcal{R}(\widehat{f}) = \int \ell(\widehat{f}(\mathbf{x}), y)\, dP(\mathbf{x}, y) = \mathbb{E}[\ell(f(X), Y)|D_n] \qquad = \mathbb{E}_{X,Y}\left[\ell(\widehat{f}(X), Y)\right]$$

$$\mathbb{E}[\mathcal{R}(\widehat{f})] = \int \ell(\widehat{f}(\mathbf{x}), y)\, dP(\mathbf{x}, y) \prod_{i=1}^{n} dP(\mathbf{x}_i, y_i) = \mathbb{E}[\ell(\widehat{f}(X), Y)] \quad = \mathbb{E}_{X,Y,D_n}\left[\ell(\widehat{f}(X), Y)\right]$$

$$\mathcal{R}(a|\mathbf{x}) = \mathbb{E}[\ell(\quad a \quad, Y)|X = \mathbf{x}] \qquad = \mathbb{E}_{X,Y}[\ell(\quad a \quad, Y)|X = \mathbf{x}] \neq \mathbb{E}_{Y}[\ell(a, Y)]$$

$$\mathcal{R}(f(X)|X) = \mathbb{E}[\ell(f(X), Y)|X] \qquad = \mathbb{E}_{X,Y}[\ell(f(X), Y)|X] \neq \mathbb{E}_{Y}[\ell(f(X), Y)] \quad \mathbf{???!}$$

$$\mathcal{R}(f(\mathbf{x})|\mathbf{x}) = \mathbb{E}[\ell(f(\mathbf{x}), Y)|X = \mathbf{x}] \qquad = \mathbb{E}_{X,Y}[\ell(f(\mathbf{x}), Y)|X = \mathbf{x}]$$

$$\mathcal{R}(\widehat{f}(\mathbf{x})|\mathbf{x}) = \mathbb{E}[\ell(\widehat{f}(\mathbf{x}), Y)|X = \mathbf{x}, D_n] \quad = \mathbb{E}_{X,Y}[\ell(\widehat{f}(\mathbf{x}), Y)|X = \mathbf{x}]$$