Lema (Métodos de ensamble)

La computadora aprende de una tarea T (var. respuesta), generando un resultado óptimo a partir de uan función de costos L a través de de experiencia E (var. explicativas). Así, el mejor modelo para E y T es aquel que minimiza la medida $\mathbb{E}$ de L. Esto es:

$$\mathbb{E}[L](y) \ = \int\int L(t, y(X))Pr(t, X) \ = \int\int (t - y(X))^2 Pr(t, X)$$

Asumir que $Pr(t|X)$ es conocida. Por demostrar:

$$\mathbb{E}[L](y) \ = \int \left((y(X) - \mathbb{E}[t|X])^2 + Var[t|X]\right) Pr(X)dX$$

Demostración

Notar que:

$$\mathbb{E}[L](y) = \int\int (t - y(X))^2 Pr(t, X) = \int\int (y(X) - \mathbb{E}[t|X] + \mathbb{E}[t|X] - t)^2 Pr(t, X) dX dt.$$

Desarrollando un poco más y recordando que $Pr(t|X) = \frac{Pr(t,X)}{Pr(X)}$:

$$\int\int (y(X) - \mathbb{E}[t|X] + \mathbb{E}[t|X] - t)^2 Pr(t, X) dX dt =$$

$$\int Pr(X) \int \left((y(X) - \mathbb{E}[t|X])^2 + (\mathbb{E}[t|X] - t)^2 - 2\mathbb{E}[t|X]^2 + 2y(X)\mathbb{E}[t|X] - 2ty(X) + 2t\mathbb{E}[t|X] \right) Pr$$

$$\int Pr(X) \left(\int (y(X) - \mathbb{E}[t|X])^2 Pr(t|X) dt + \int (\mathbb{E}[t|X] - t)^2 Pr(t|X) dt - \int 2\mathbb{E}[t|X]^2 Pr(t|X) dt + \int \right.$$

$$\left. \int 2ty(X) Pr(t|X) dt + \int 2t\mathbb{E}[t|X] Pr(t|X) dt \right) dX.$$

Notar que, por definición de esperanza condicional, la función $\mathbb{E}[t|X] = \int_{\Omega_t} t Pr(t|X)$ ya es constante respecto a t. Asimismo, la integral $\int (\mathbb{E}[t|X] - t)^2 Pr(t|X) dt = \mathbb{E}[(t - \mathbb{E}[t|X])^2 | x]$ es la varianza condicional de t dado X denotada por la expresión $Var[t|X]$.

Usando lo anterior, y recordando que $y(X) = w^T X$ (la predicción del modelo) no depende de t, al integrar sobre t se obtiene que:

1. $-\int 2\mathbb{E}[t|X]^2 Pr(t|X) dt = -2\mathbb{E}[t|X]^2 \int Pr(t|X) dt = -2\mathbb{E}[t|X]^2 \cdot 1 = -2\mathbb{E}[t|X]^2$
2. $\int 2y(X)\mathbb{E}[t|X] Pr(t|X) dt = 2y(X)\mathbb{E}[t|X] \int Pr(t|X) dt = 2y(X)\mathbb{E}[t|X]$
3. $-\int 2ty(X) Pr(t|X) dt = -2y(X) \int t Pr(t|X) dt = -2y(X)\mathbb{E}[t|X]$
4. $\int 2t\mathbb{E}[t|X] Pr(t|X) dt = 2\mathbb{E}[t|X] \int t Pr(t|X) dt = 2\mathbb{E}[t|X]^2$
5. $\int (y(X) - \mathbb{E}[t|X])^2 Pr(t|X) dt = (y(X) - \mathbb{E}[t|X])^2 \int Pr(t|X) dt = (y(X) - \mathbb{E}[t|X])^2$

Así, sumando las integrales 1. a 5. tenemos que:

$$\mathbb{E}[L](y) = \int Pr(X) \left((y(X) - \mathbb{E}[t|X])^2 + Var[t|X] - 2\mathbb{E}[t|X]^2 + 2\mathbb{E}[t|X]^2 + 2y(X)\mathbb{E}[t|X] - 2y(X)\right.$$

$$= \int \left((y(X) - \mathbb{E}[t|X])^2 + Var[t|X] \right) Pr(X) dX \blacksquare$$

Como corolario del lema anterior, dado que $(y(X) - \mathbb{E}[t|x])^2 \geq 0,\ Var[t|X] \geq 0,\ Pr(X) \geq 0,$ el modelo que minimiza la medida de la función de costos L es $y^* = \mathbb{E}[t|x]$ (ya que $(y^*(X) - \mathbb{E}[t|x])^2 = 0$).