Definitions of conditional expectations & distribution functions :-

$$E(XY) = \sum_i \sum_j x_i y_j p(x_i, x_j) \quad - (1)$$

$$E(x^2) = \sum_i \sum_j x_i^2 p(x_i, y_j) \quad - (2)$$

Joint PDF, $\sum_i \sum_j p(x_i, y_j) = 1 \quad - (3)$

(probability of joint pdf = 1,)

Differentiating (1) & (2) w.r.t ∂x_i,

$$\frac{\partial E(x^2)}{\partial x_i} = 2x_i \sum_j y_j p(x_i, x_j)$$

$$\frac{\partial E(XY)}{\partial x_i} = \sum_j y_j p(x_i, y_j).$$

The differentiating of the term $p(x_i, y_j)$ leads to 0 as its summation is a constant.

$$\therefore, \quad E(Y_i) = \sum_j \theta(Y_j) \times p(x_i, Y_j)$$

$$= \frac{\sum_j \theta(Y_j) \times p(x_i, Y_j)}{p(x_i, Y_j)}$$

Minimizing e^2,

$$e^2 = E\{[\theta(Y) - \phi(X)]^2\} + \lambda\,(1 - (E(\theta^2(Y))))$$

$$= E(\theta^2(Y)) + 2E[\theta(Y) \times \phi X)] + E[\phi^2(X)]$$
$$+ \lambda - \lambda(E(\theta^2(Y)))$$

Expanding the terms, wrt E,

$$= \Sigma_i \Sigma_j \theta^2(Y_j)\, p(x_i, x_j) + 2\Sigma_i \Sigma_j \phi(x_i)\theta(Y_j) \times p(x_i, Y_j)$$
$$+ \Sigma_i \Sigma_j \phi^2(X_i) p(X_i, Y_j) + \lambda - \lambda[\Sigma_i \Sigma_i \theta^2(Y_j) p(x_i, x_j)]$$

At the minima, the partial derivatives of independent variables are 0.

$$\therefore \frac{\partial e^2(\theta, \phi, \lambda)}{\partial \lambda} = 0$$

$$\frac{\partial e^2(\theta, \phi, \lambda)}{\partial \phi} = 0$$

$$\frac{\partial e^2(\theta, \phi, \lambda)}{\partial \theta} = 0$$

$$\frac{\partial e^2(\theta,\phi,\lambda)}{\partial\phi} = 2\sum_j \theta(Y_j) \times p(x_i,Y_j) + 2\phi(x_i)\sum_j p(x_i,Y_j) = 0,$$

Solving for, $\phi(x_i)$,

$$\phi(x_i) = \frac{\sum_j \theta(Y_j) \times p(x_i,Y_j)}{p(x_i,Y_j)} = E(\theta(Y)|X_i)$$

$$\phi(x) = E(\theta(Y)|X)$$

$\Rightarrow$ To minimise wrt $\theta(Y_j)$

differentiate wrt λ,

$$\frac{\partial e^2(\theta,\phi,\lambda)}{\partial\lambda} = 1-[E(\theta^2(Y))] = 0$$

$\Rightarrow$ Differentiate wrt θ,

$$\frac{\partial e^2(\theta,\phi,\lambda)}{\partial\theta} = 2\theta(Y_j)\sum_i p(x_i,Y_j) - 2\sum_i \phi(x_i)p(x_i,Y_j) - 2\lambda\theta(Y_j)\sum_i p(x_i,Y_j) = (1-\lambda)\theta(Y_j)\sum_i p(x_i,Y_j) - \sum_i \phi(x_i)p(x_i,Y_j) = 0$$

$$\theta(Y_j) = \frac{1}{1-\lambda} * \frac{\sum_i \phi(x_i) \times p(x_i, x_j)}{\sum_i p(x_i, x_j)} = \frac{1}{1-\lambda} E(\phi(x_i) | Y_j)$$

Substitutin,

$$1-(E(\theta^2(Y))) = 1 - E\left[\left(\frac{1}{1-\lambda} E(\phi(x_i)|Y_j)\right)^2\right]$$

$$= 1 - \left(\frac{1}{1-\lambda}\right)^2 \times \sum_i \sum_j ((E(\phi(x_i)|Y_j))^2 \times p(x_i, Y_j))$$

$$= 1 - \left(\frac{1}{1-\lambda}\right)^2 \times \left[(E(\phi(x_i)|Y_i))^2 \sum_i p(x_i, Y_j)\right)$$

$$= 1 - \left(\frac{1}{1-\lambda}\right)^2 \times \sum_j ((E(\phi(x_i)|Y_j))^2 p(x_i, Y_j)) = 0$$

considering a ducute distrib.

$$1-\lambda = \sqrt{\sum_j (E(\phi(x_i)|Y_j))^2 p(x_i, Y_j)}) = abs(E(\phi(x)|Y)$$

sme, $\theta(Y) = \frac{1}{1-\lambda} E(Y_j) = \frac{E(\phi(x)Y)}{||E(\phi(x)|Y)||}$

$$\theta(Y) = \frac{E(\phi(x)|Y)}{||E(\phi(x)|Y||}$$