Summary of EE226a: MMSE and LLSE

Jean Walrand

I. Summary

Here are the key ideas and results:

- The MMSE of $\mathbf{X}$ given $\mathbf{Y}$ is $E[\mathbf{X}|\mathbf{Y}]$.
- The LLSE of $\mathbf{X}$ given $\mathbf{Y}$ is $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}}K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y}))$ if $K_{\mathbf{Y}}$ is nonsingular.
- Theorem 3 states the properties of the LLSE.
- The linear regression approximates the LLSE when the samples are realizations of i.i.d. random pairs (X_m, Y_m).

II. Estimation: Formulation

The estimation problem is a generalized version of the Bayesian decision problem where the set of values of $\mathbf{X}$ can be $\Re^n$. In applications, one is given a *source model* $f_{\mathbf{X}}$ and a *channel model* $f_{\mathbf{Y}|\mathbf{X}}$ that together define $f_{\mathbf{X},\mathbf{Y}}$. One may have to estimate $f_{\mathbf{Y}|\mathbf{X}}$ by using a *training sequence*, i.e., by selecting the values of $\mathbf{X}$ and observing the channel outputs. Alternatively, one may be able to observe a sequence of values of $(\mathbf{X}, \mathbf{Y})$ and use them to estimate $f_{\mathbf{X},\mathbf{Y}}$.

Definition 1: Estimation Problems

One is given the joint distribution of $(\mathbf{X}, \mathbf{Y})$. The estimation problem is to calculate $\mathbf{Z} = g(\mathbf{Y})$ to minimize $E(c(\mathbf{X}, \mathbf{Z}))$ for a given function $c(\cdot, \cdot)$. The random variable $\mathbf{Z} = g(\mathbf{Y})$ that minimizes $E(||\mathbf{X} - \mathbf{Z}||^2)$ is the *minimum mean squares estimator* (MMSE) of $\mathbf{X}$ given $\mathbf{Y}$. The random variable $\mathbf{Z} = A\mathbf{Y} + \mathbf{b}$ that minimizes $E(||\mathbf{X} - \mathbf{Z}||^2)$ is called the *linear least squares estimator* (LLSE) of $\mathbf{X}$ given $\mathbf{Y}$; we designate it by $\mathbf{Z} = L[\mathbf{X}|\mathbf{Y}]$.

III. MMSE

Here is the central result about minimum mean squares estimation. You have seen this before, but we recall the proof of that important result.

Theorem 1: MMSE

The MMSE of $\mathbf{X}$ given $\mathbf{Y}$ is $E[\mathbf{X}|\mathbf{Y}]$.

Proof: You should recall the definition of $E[\mathbf{X}|\mathbf{Y}]$, a random variable that has the property

$$E[(\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])h_1(\mathbf{Y})] = 0, \forall h_1(\cdot), \tag{1}$$

or equivalently

$$E[h_2(\mathbf{Y})(\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])] = 0, \forall h_2(\cdot). \tag{2}$$

By $E(\mathbf{X}) = E[E[\mathbf{X}|\mathbf{Y}]]$, the interpretation is that $\mathbf{X} - E[\mathbf{X}|\mathbf{Y}] \perp h(\mathbf{Y})$ for all $h(\cdot)$. By Pythagoras, one then expect $E[\mathbf{X}|\mathbf{Y}]$ to be the function of $\mathbf{Y}$ that is closest to $\mathbf{X}$, as illustrated in Figure 1.

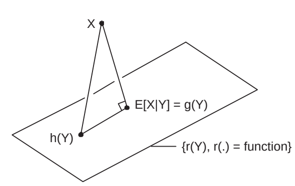

Fig. 1. The conditional expectation as a projection.

Formally, let $h(\mathbf{Y})$ be an arbitrary function. Then

$$\begin{aligned} E(||\mathbf{X} - h(\mathbf{Y})||^2) &= E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}] + E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y})||^2) \\ &= E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]||^2) + 2E((E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y}))^T(\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])) + E(||E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y})||^2) \\ &= E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]||^2) + E(||E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y})||^2) \geq E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]||^2). \end{aligned}$$

In this derivation, the third identity follows from the fact that the cross-term vanishes in view of (2). Note also that this derivation corresponds to the fact that the triangle $\{\mathbf{X}, E[\mathbf{X}|\mathbf{Y}], h(\mathbf{Y})\}$ is a right triangle with hypothenuse $(\mathbf{X}, h(\mathbf{Y}))$. ∎

Example 1: You recall that if $(\mathbf{X}, \mathbf{Y})$ are jointly Gaussian with $K_{\mathbf{Y}}$ nonsingular, then

$$E[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y})).$$

You also recall what happens when $K_{\mathbf{Y}}$ is singular. The key is to find A so that $K_{\mathbf{X},\mathbf{Y}} = AK_{\mathbf{Y}} = AQ\Lambda Q^T$. By writing

$$Q = [Q_1|Q_2], \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}$$

where Λ_1 corresponds the part of nonzero eigenvalues of $K_{\mathbf{Y}}$, one finds

$$K_{\mathbf{X},\mathbf{Y}} = AQ\Lambda Q^T = A[Q_1|Q_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = AQ_1\Lambda_1 Q_1^T,$$

so that $K_{\mathbf{X},\mathbf{Y}} = AK_{\mathbf{Y}}$ with $A = K_{\mathbf{X},\mathbf{Y}} Q_1 \Lambda_1^{-1} Q_1^T$.

IV. LLSE

Theorem 2: LLSE

$L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y}))$ if $K_{\mathbf{Y}}$ is nonsingular. $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}} Q_1\Lambda_1^{-1}Q_1^T(\mathbf{Y} - E(\mathbf{Y}))$ if $K_{\mathbf{Y}}$ is singular.

Proof: $\mathbf{Z} = L[\mathbf{X}|\mathbf{Y}] = A\mathbf{Y} + \mathbf{b}$ satisfies $\mathbf{X} - \mathbf{Z} \perp B\mathbf{Y} + \mathbf{d}$ for any B and $\mathbf{d}$ with $E(\mathbf{X}) = E[L[\mathbf{X}|\mathbf{Y}]]$. If we consider any $\mathbf{Z}' = C\mathbf{Y} + \mathbf{c}$, then $E((\mathbf{Z} - \mathbf{Z}')^T(\mathbf{X} - \mathbf{Z})) = 0$ since $\mathbf{Z} - \mathbf{Z}' = A\mathbf{Y} + \mathbf{b} - C\mathbf{Y} - \mathbf{c} = (A - C)\mathbf{Y} + (\mathbf{b} - \mathbf{c}) = B\mathbf{Y} + \mathbf{d}$. It follows that

$$\begin{aligned} E(||\mathbf{X} - \mathbf{Z}'||^2) &= E(||\mathbf{X} - \mathbf{Z} + \mathbf{Z} - \mathbf{Z}'||^2) \\ &= E(||\mathbf{X} - \mathbf{Z}||^2) + 2E((\mathbf{Z} - \mathbf{Z}')^T(\mathbf{X} - \mathbf{Z})) + E(||\mathbf{Z} - \mathbf{Z}'||^2) \\ &= E(||\mathbf{X} - \mathbf{Z}||^2) + E(||\mathbf{Z} - \mathbf{Z}'||^2) \geq E(||\mathbf{X} - \mathbf{Z}||^2). \end{aligned}$$

Figure 2 illustrates this calculation. It shows that $L[\mathbf{X}|\mathbf{Y}]$ is the projection of $\mathbf{X}$ on the set of linear functions of $\mathbf{Y}$. The projection $\mathbf{Z}$ is characterized by the property that $\mathbf{X} - \mathbf{Z}$ is orthogonal to all the linear functions $B\mathbf{Y} + \mathbf{d}$. This property holds if and only if (from (1))

$$E(\mathbf{X} - \mathbf{Z}) = 0 \text{ and } E((\mathbf{X} - \mathbf{Z})\mathbf{Y}^T) = 0,$$

which are equivalent to $E(\mathbf{X} - \mathbf{Z}) = 0$ and $\text{cov}(\mathbf{X} - \mathbf{Z}, \mathbf{Y}) = 0$. The comparison between $L[\mathbf{X}|\mathbf{Y}]$ and $E[\mathbf{X}|\mathbf{Y}]$ is shown in Figure 3. ■

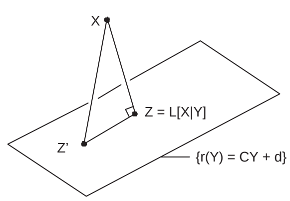

Fig. 2. The LLSE as a projection.

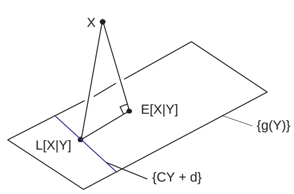

Fig. 3. LLSE vs. MMSE.

The following result indicates some properties of the LLSE. They are easy to prove.

Theorem 3: Properties of LLSE

(a) $L[A\mathbf{X}_1 + B\mathbf{X}_2|\mathbf{Y}] = AL[\mathbf{X}_1|\mathbf{Y}] + BL[\mathbf{X}_2|\mathbf{Y}]$.

(b) $L[L[\mathbf{X}|\mathbf{Y}, \mathbf{Z}] \mid \mathbf{Y}] = L[\mathbf{X}|\mathbf{Y}]$.

(c) If $\mathbf{X} \perp \mathbf{Y}$, then $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X})$.

(d) Assume that X and Z are conditionally independent given Y. Then, in general, $L[X|Y, Z] \neq L[X|Y]$.

Proof: (b): It suffices to show that

$$E(L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] - L[\mathbf{X}|\mathbf{Y}]) = 0 \text{ and } E((L[\mathbf{X}|\mathbf{Y}] - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}])\mathbf{Y}^T) = 0.$$

We already have $E(X - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}]) = 0$ and $E(X - L[\mathbf{X}|\mathbf{Y}]) = 0$, which implies $E(L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] - L[\mathbf{X}|\mathbf{Y}]) = 0$. Moreover, from $E((\mathbf{X} - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}])\mathbf{Y}^T) = 0$ and $E((\mathbf{X} - L[\mathbf{X}|\mathbf{Y}])\mathbf{Y}^T) = 0$, it is obtained $E((L[\mathbf{X}|\mathbf{Y}] - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}])\mathbf{Y}^T) = 0$.

(d): Here is a trivial example. Assume that $X = Z = Y^{1/2}$ where X is $U[0,1]$. Then $L[X|Y,Z] = Z$. However, since $Y = X^2$, $Y^2 = X^4$, and $XY = X^3$, we find

$$\begin{aligned} L[X|Y] &= E(X) + \frac{E(XY) - E(X)E(Y)}{E(Y^2) - E(Y)^2}(Y - E(Y)) \\ &= \frac{1}{2} + \left(\frac{1}{4} - \frac{1}{2}\cdot\frac{1}{3}\right)\left(\frac{1}{5} - \frac{1}{9}\right)^{-1}\left(Y - \frac{1}{3}\right) \\ &= \frac{3}{16} + \frac{15}{16}Y. \end{aligned}$$

We see that $L[X|Y,Z] \neq L[X|Y]$ because

$$Z \neq \frac{3}{16} + \frac{15}{16}Y = \frac{3}{16} + \frac{15}{16}Z^2.$$

∎

Figure 4 illustrates property (b) which is called the *smoothing property* of LLSE.

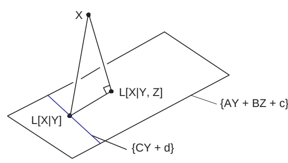

Fig. 4. Smoothing property of LLSE.

V. Examples

We illustrate the previous ideas on a few examples.

Example 2: Assume that U, V, W are independent and $U[0,1]$. Calculate $L[(U+V)^2 \mid V^2 + W^2]$.

Solution: Let $X = (U+V)^2$ and $Y = V^2 + W^2$. Then

$$K_{X,Y} = \text{cov}(U^2 + 2UV + V^2, V^2 + W^2) = 2\text{cov}(UV, V^2) + \text{cov}(V^2, V^2).$$

Now,

$$\text{cov}(UV, V^2) = E(UV^3) - E(UV)E(V^2) = \frac{1}{8} - \frac{1}{12} = \frac{1}{24}$$

and

$$\text{cov}(V^2, V^2) = E(V^4) - E(V^2)E(V^2) = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}.$$

Hence,

$$K_{X,Y} = \frac{1}{12} + \frac{4}{45} = \frac{31}{180}.$$

We conclude that

$$L[(U+V)^2 \mid V^2 + W^2] = E((U+V)^2) + K_{X,Y}K_Y^{-1}(Y - E(Y)) = \frac{7}{6} + \frac{31}{32}\left(Y - \frac{2}{3}\right).$$

Example 3: Let U, V, W be as in the previous example. Calculate $L[\cos(U+V)|V+W]$.

Solution: Let $X = \cos(U+V)$ and $Y = V + W$. We find

$$K_{X,Y} = E(XY) - E(X)E(Y).$$

Now,

$$E(XY) = E(V\cos(U+V)) + \frac{1}{2}E(X).$$

Also,

$$\begin{aligned} E(V\cos(U+V)) &= \int_0^1 \int_0^1 v\cos(u+v)dudv = \int_0^1 [v\sin(u+v)]_0^1 dv \\ &= \int_0^1 (v\sin(v+1) - v\sin(v))dv = -\int_0^1 v \cdot d\cos(v+1) + \int_0^1 v \cdot d\cos(v) \\ &= -[v\cos(v+1)]_0^1 + \int_0^1 \cos(v+1)dv + [v\cos(v)]_0^1 - \int_0^1 \cos(v)dv \\ &= -\cos(2) + [\sin(v+1)]_0^1 + \cos(1) - [\sin(v)]_0^1 \\ &= -\cos(2) + \sin(2) - \sin(1) + \cos(1) - \sin(1). \end{aligned}$$

Moreover,

$$\begin{aligned} E(X) &= \int_0^1 \int_0^1 \cos(u+v)dudv = \int_0^1 [\sin(u+v)]_0^1 du = \int_0^1 (\sin(u+1) - \sin(u))du \\ &= -[\cos(u+1)]_0^1 + [\cos(u)]_0^1 = -\cos(2) + \cos(1) + \cos(1) - \cos(0) = -\cos(2) + 2\cos(1) - 1. \end{aligned}$$

In addition,

$$E(Y) = E(V) + E(U) = 1.$$

and

$$K_Y = \frac{1}{6}.$$

VI. Linear Regression vs. LLSE

We discuss the connection between the familiar *linear regression* procedure and the LLSE.

One is given a set of n pairs of numbers $\{(x_m, y_m), m = 1, \ldots, n\}$, as shown in Figure 5. One draws a line through the points. That line approximates the values y_m by $z_m = \alpha x_m + \beta$. The line is chosen to minimize

$$\sum_{m=1}^{n} (z_m - y_m)^2 = \sum_{m=1}^{n} (\alpha x_m + \beta - y_m)^2. \tag{3}$$

That is, the linear regression is the linear approximation that minimizes the sum of the squared errors. Note that there is no probabilistic framework in this procedure.

To find the values of α and β that minimize the sum of the squared errors, one differentiates (3) with respect to α and β and sets the derivatives to zero. One finds

$$\sum_{m=1}^{n} (\alpha x_m + \beta - y_m) = 0 \text{ and } \sum_{m=1}^{n} x_m(\alpha x_m + \beta - y_m) = 0.$$

Solving these equations, we find

$$\alpha = \frac{A(xy) - A(x)A(y)}{A(x^2) - A(x)^2} \text{ and } \beta = A(y) - \alpha A(x).$$

In these expressions, we used the following notation:

$$A(y) := \frac{1}{n}\sum_{m=1}^{n} y_m, A(x) := \frac{1}{n}\sum_{m=1}^{n} x_m, A(xy) := \frac{1}{n}\sum_{m=1}^{n} x_m y_m, \text{ and } A(x^2) := \frac{1}{n}\sum_{m=1}^{n} x_m^2.$$

Thus, the point y_m is approximated by

$$z_m = A(y) + \frac{A(xy) - A(x)A(y)}{A(x^2) - A(x)^2}(x_m - A(x)).$$

Note that if the pairs (x_m, y_m) are realizations of i.i.d. random variables $(X_m, Y_m) =_D (X, Y)$ with finite variances, then, as $n \to \infty$, $A(x) \to E(X)$, $A(x^2) \to E(X^2)$, $A(y) \to E(Y)$, and $A(xy) \to E(XY)$. Consequently, if n is large, we see that

$$z_m \approx A(Y) + \frac{\text{cov}(X,Y)}{\text{var}(X)}(x_m - E(X)) = L[Y|X = x_m].$$

See [2] for examples.

References

[1] R. G. Gallager: *Stochastic Processes: A Conceptual Approach*. August 20, 2001.
[2] D. Freedman: *Statistical Models: Theory and Practice*. Cambridge University Press, 2005.
[3] J Walrand: *Lecture Notes on Probability and Random Processes*, On-Line Book, August 2004.

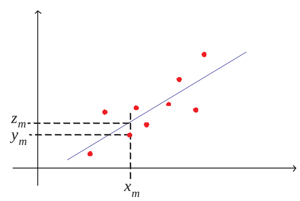

Fig. 5. Linear Regression.