Mejor Predicción Lineal: modelo lineal con ruido identificado

Presentación en vídeo: http://personales.upv.es/asala/YT/V/vcinv1.html

Objetivos: comprender la relación entre las fórmulas de "*mejor predicción lineal*" basadas en la matriz VC de dos variables aleatorias (a, b) y los modelos lineales con ruido $a = \theta b + \epsilon$ que se pueden identificar a partir de dichas fórmulas.

Tabla de Contenidos

Preliminares, conceptos previos

Mejor predicción lineal

Dadas dos variables aleatorias a y b, suponemos media cero por simplificar (si no, se cambiaría todo a "incrementos sobre la media").

Si su matriz varianzas-covarianzas, simétrica, es:

$\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$, con $\Sigma_{ab} = \Sigma_{ba}^T$ (si fuera matriz, caso multivariable), entonces mejor pred. lineal de a dado b es:

$\hat{a} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot b$, con una varianza del error de predicción $a - \hat{a}$ dada por: $\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}^T$.

En caso de distribución normal, conocida varianza y media se conoce todo, con lo que la probabilidad condicional $a|b$ es una distribución $N(\hat{a}, \Sigma_{e,a})$.

Matriz de varianzas-covarianzas asociada a un modelo lineal con ruido

Dado un modelo $a = \theta b + \epsilon$, con $b \sim N(0, \Sigma_{bb})$, $\epsilon \sim N(0, \Sigma_e)$, siendo ε independiente de b, entonces:

$$\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^T] = E[\theta b b^T \theta^T + \epsilon b^T \theta^T + \theta b \epsilon^T + \epsilon \epsilon^T]$$
$$= \theta E[bb^T]\theta^T + E[\epsilon b^T]\theta^T + \theta E[b\epsilon^T] + E[\epsilon\epsilon^T] = \theta \Sigma_{bb} \theta^T + \Sigma_e$$

$$\Sigma_{ab} = E[(\theta b + \epsilon) b^T] = \theta E[bb^T] + E[\epsilon b^T] = \theta \Sigma_{bb}$$

Con lo que la matriz VC conjunta de a y b sería:

$$\Sigma := \begin{pmatrix} \theta \Sigma_{bb} \theta^T + \Sigma_e & \theta \Sigma_{bb} \\ \Sigma_{bb} \theta^T & \Sigma_{bb} \end{pmatrix}$$

El objetivo de este material es comprender la relación entre "mejor predicción lineal" y "modelo lineal con ruido" identificado.

Modelos identificados a partir de una matriz VC

Consideremos $\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$.

Modelo identificado de predicción de a dado b

Si llamamos $\theta := \Sigma_{ab} \Sigma_{bb}^{-1}$, entonces la predicción es $\hat{a} = \theta \cdot b$, con una varianza de error de predicción $\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \cdot \Sigma_{bb} \cdot \Sigma_{bb}^{-1} \Sigma_{ab}^T = \Sigma_{aa} - \theta \Sigma_{bb} \theta^T$.

Si ahora suponemos un modelo:

$a = \theta \cdot b + \epsilon$, siendo $\epsilon \sim N(0, \Sigma_{e,a})$ estadísticamente independiente de a y b, y b una variable aleatoria con media cero y varianza Σ_{bb}

entonces tendríamos $\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^T] = \theta \Sigma_{bb} \theta^T + \Sigma_{e,a}$ que devuelve la expresión de arriba.

También $\Sigma_{ab} = \theta \Sigma_{bb} = \Sigma_{ab} \Sigma_{bb}^{-1} \cdot \Sigma_{bb}$ devuelve la covarianza inicial.

Por tanto, podríamos decir que el modelo lineal con ruido $a = \theta b + \epsilon$, con la varianza de b siendo Σ_{bb} y la varianza de ϵ siendo $\Sigma_{e,a}$ "explica" la matriz de varianzas covarianzas entre a y b.

```
Sigma=[4 3;3 8];
```

```
canonicas=0;
if(canonicas) %escalar a dt = 1 todas las variables
    dsvt=sqrt(diag(Sigma)); %desv. típicas de cada variable
    EscM=inv(diag(dsvt));
    Sigma=EscM*Sigma*EscM'
end
eig(Sigma) %debe ser def+ para que tenga sentido estadístico
```

```
ans = 2×1
    2.3944
    9.6056
```

```
theta=Sigma(1,2)*inv(Sigma(2,2))
```

```
theta = 0.3750
```

```
vza_ea=Sigma(1,1)-Sigma(1,2)*inv(Sigma(2,2))*Sigma(2,1)
```

```
vza_ea = 2.8750
```

```
desvtip_ea=sqrt(vza_ea)
```

```
desvtip_ea = 1.6956
```

Por tanto, la covarianza entre a y b podría ser explicada por $a = \theta \cdot b + \epsilon$.

Modelo identificado de predicción de b dado a

Obviamente, podemos cambiar nombres y plantear un modelo

$b = \eta a + \varepsilon$, con $\eta = \Sigma_{ba}\Sigma_{aa}^{-1}$ (fórmula mejor pred. lineal), y ruido $\varepsilon \sim N(0, \Sigma_{e,b})$, independiente de a y de b, con $\Sigma_{e,b} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ba}^T = \Sigma_{bb} - \eta\Sigma_{aa}\eta^T$

Haciendo operaciones análogas, el modelo $b = \eta a + \varepsilon$ con a de varianza Σ_{aa} y ε de varianza $\sigma_{e,b}$ explica también la matriz de varianzas-covarianzas entre a y b.

```
eta=Sigma(2,1)*inv(Sigma(1,1))
```

```
eta = 0.7500
```

```
vza_eb=Sigma(2,2)-Sigma(2,1)*inv(Sigma(1,1))*Sigma(1,2)
```

```
vza_eb = 5.7500
```

```
desvtip_eb=sqrt(vza_eb)
```

```
desvtip_eb = 2.3979
```

Relación entre ambos modelos

Si el modelo fuera "determinista" y el número de variables en a y b fuera idéntico, entonces $a = \theta b + \epsilon$ implicaría $b = \theta^{-1}a - \theta^{-1}\epsilon$, pero NO es el caso

```
inv(theta) %no es "eta"
```

```
ans = 2.6667
```

```
inv(theta)*desvtip_ea %no es desvtip_eb
```

```
ans = 4.5216
```

Dibujemos ambos modelos, junto con cierta informacion adicional para comprender mejor...

- Algunas muestras

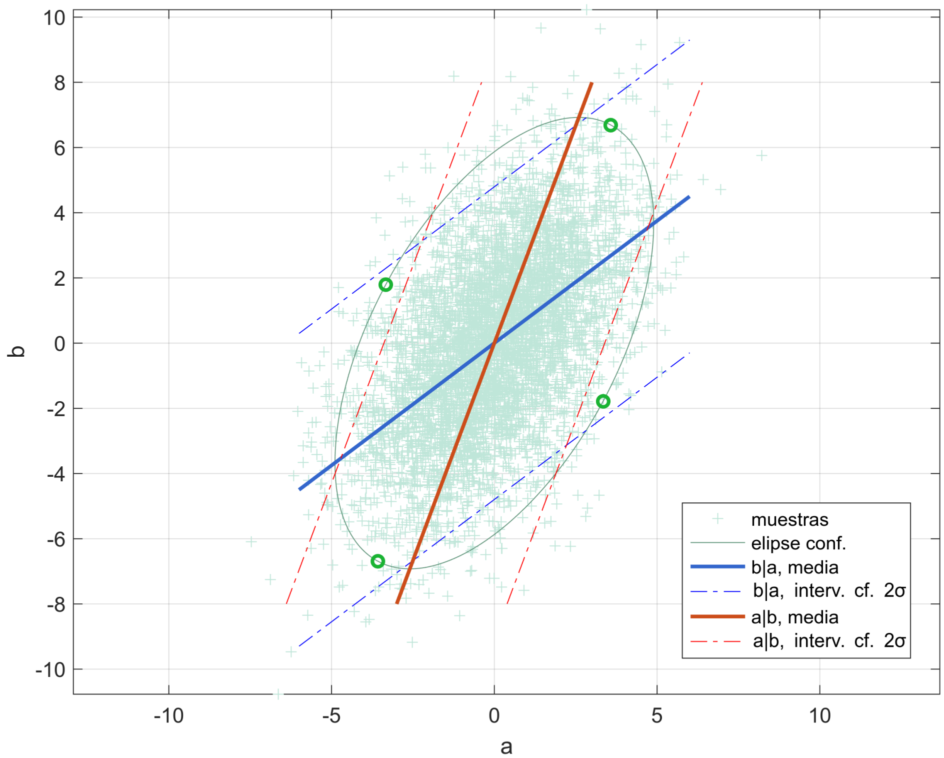
```
R = mvnrnd([0 0],Sigma,5000);
plot(R(:,1),R(:,2),'+',Color=[0.75 0.9 0.85]) %muestras
hold on
```

- Elipsoide de confianza 95% (caso distr. normal)

```
[V,D]=eig(Sigma);
alpha=(0:360)*2*pi/360;
circulo=[sin(alpha); cos(alpha)];
ell=V*sqrt(D)*sqrt(chi2inv(0.95,2))*circulo;
plot(ell(1,:),ell(2,:),Color=[.4 .6 .5])
%los ejes de la elipse...
L=round((length(alpha)-1)/4);
plot(ell(1,1:L:end),ell(2,1:L:end),'o',Color=[.1 .7 .2],LineWidth=2)
```

- Modelo de "a dado b", y de "b dado a" (media e intervalos de confianza $\pm 2\sigma$)

```
range_a=[-6 6];
range_b=[-8 8];
plot(range_a,eta*range_a,LineWidth=2,Color=[0.2 0.4 0.8]), %hold on
plot(range_a,eta*range_a+2*desvtip_eb,'-.b')
plot(range_a,eta*range_a-2*desvtip_eb,'-.b')
plot(theta*range_b,range_b,LineWidth=2,Color=[.8 .3 .1]),
plot(theta*range_b-2*desvtip_ea,range_b,'-.r')
plot(theta*range_b+2*desvtip_ea,range_b,'-.r')
hold off, grid on, axis equal
legend("muestras","elipse conf.","","b|a, media","b|a, interv. cf. 2\sigma","","a|b, me
xlabel("a"),ylabel("b")
```



*Observa la línea "b|a" cortando el elipsoide donde la tangente es "vertical", y la línea "a|b" donde la tangente es "horizontal". El modelo "intermedio" dado por el eje mayor de la elipse es el modelo de "mínimos cuadrados totales" o "componentes principales", fuera de los objetivos de este material.