

Mejor predicción lineal. Ejemplo 2 variables.

© 2018, Antonio Sala Piqueras, Universitat Politècnica de València. Todos los derechos reservados.

Presentación en vídeo en personales.upv.es/asala/YT/V/prelim.html

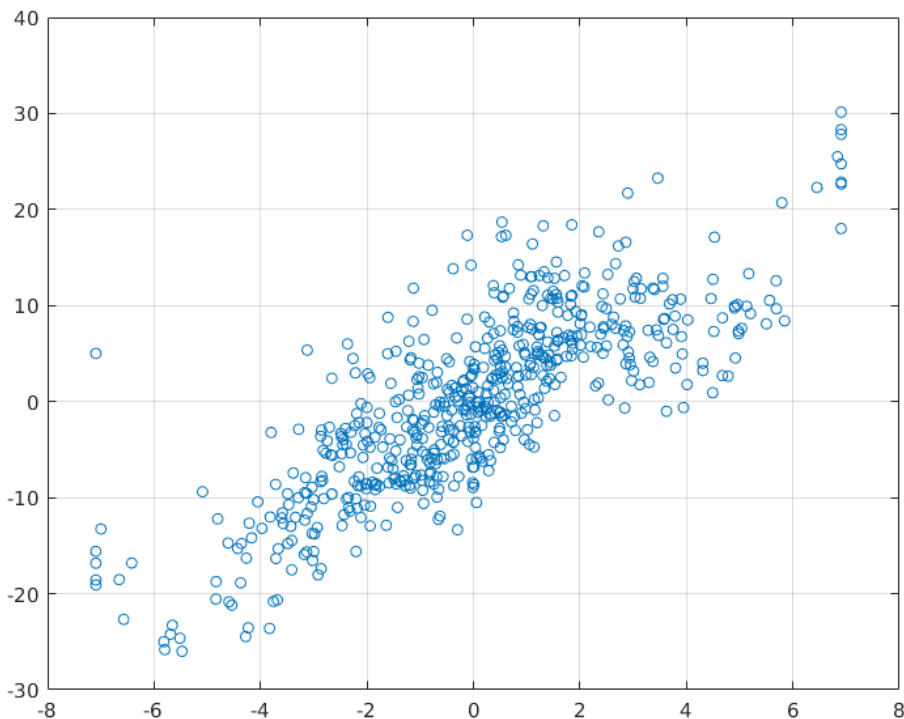
Este código ejecutó sin errores en Matlab R2018b.

Tabla de Contenidos

Generación y visualización de los datos.....	1
Análisis de varianzas-covarianzas.....	2
Predicción de y dado x.....	2
Predicción de x dado y.....	3
Gráfica de resultados.....	3

Generación y visualización de los datos

```
Nmuestras=600;  
[x,y,ysecreto]=formulasecreta(Nmuestras);  
  
xmedio=mean(x); ymedio=mean(y);  
x=x-xmedio;y=y-ymedio; %normalizamos a media 0  
ysecreto=ysecreto-ymedio;%este es "desconocido"  
plot(x,y,'o'), grid on
```



Análisis de varianzas-covarianzas

```
varx=cov(x)
```

```
varx = 6.4538
```

```
stdx=sqrt(varx)
```

```
stdx = 2.5404
```

```
vary=cov(y)
```

```
vary = 87.0770
```

```
stdy=sqrt(vary)
```

```
stdy = 9.3315
```

```
varxy=cov([x y])
```

```
varxy = 2x2  
      6.4538  19.1308  
      19.1308  87.0770
```

```
varx=varxy(1,1);  
vary=varxy(2,2);  
covxy=varxy(1,2);  
correlacion=covxy/sqrt(varx*vary)
```

```
correlacion = 0.8070
```

Predicción de y dado x

```
bestpredygivenx_coef=covxy/varx
```

```
bestpredygivenx_coef = 2.9643
```

```
%coincide con mínimos cuadrados:  
pinv(x)*y
```

```
ans = 2.9643
```

```
vary_posteriori=vary-covxy*inv(varx)*covxy
```

```
vary_posteriori = 30.3680
```

Comparemos la desviación típica sin información (a priori) y la desviación típica a posteriori

```
stdypriori=sqrt(vary)
```

```
stdypriori = 9.3315
```

```
stdyposteriori=sqrt(vary_posteriori)
```

```
stdyposteriori = 5.5107
```

```
porcent_NO_explicado=stdyposteriori/stdypriori
```

```
porcent_NO_explicado = 0.5905
```

```
sqrt(1-correlacion^2)
```

```
ans = 0.5905
```

Predicción de x dado y

```
bestpredxgiveny_coef=covxy/vary
```

```
bestpredxgiveny_coef = 0.2197
```

```
%coincide con mínimos cuadrados:  
pinv(y)*x
```

```
ans = 0.2197
```

```
varx_posteriori=varx-covxy*inv(vary)*covxy
```

```
varx_posteriori = 2.2508
```

Comparemos la desviación típica sin información (a priori) y la desviación típica a posteriori

```
stdxpriori=sqrt(varx)
```

```
stdxpriori = 2.5404
```

```
stdxposteriori=sqrt(varx_posteriori)
```

```
stdxposteriori = 1.5003
```

```
porcent_NO_explicadox=stdxposteriori/stdxpriori
```

```
porcent_NO_explicadox = 0.5905
```

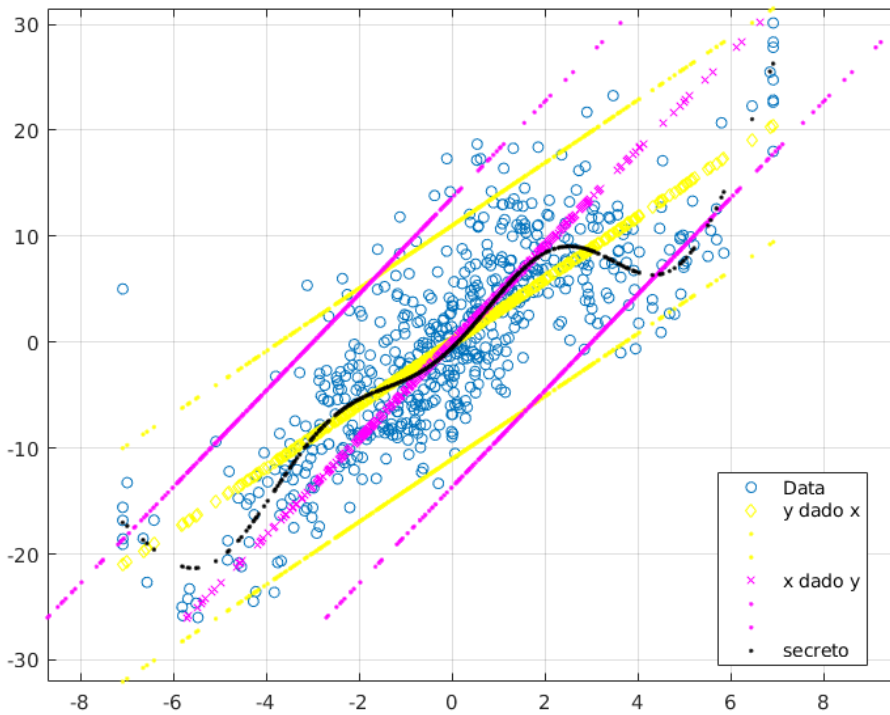
```
sqrt(1-correlacion^2)
```

```
ans = 0.5905
```

Gráfica de resultados

```
plot(x,y,'o')  
hold on  
plot(x,bestpredygivenx_coef*x,'dy')  
plot(x,bestpredygivenx_coef*x+2*stdyposteriori,'.y')  
plot(x,bestpredygivenx_coef*x-2*stdyposteriori,'.y')  
plot(bestpredxgiveny_coef*y,y,'xm')  
plot(bestpredxgiveny_coef*y+2*stdxposteriori,y,'.m')  
plot(bestpredxgiveny_coef*y-2*stdxposteriori,y,'.m')  
plot(x,ysecreto,'.k')  
hold off  
grid on
```

```
axis tight
legend('Data','y dado x','','','x dado y','','','secreto','Location','best')
```



```
function [xc,y,ysecreto]=formulasecreta(Nmuestras)
    xc=max(-7,min(7,randn(Nmuestras,1)*2+(randn(Nmuestras,1).^3)*.5));
    n=(randn(Nmuestras,1)).^2*2.5+randn(Nmuestras,1)*3.75;
    ysecreto=3*xc+sin(xc).*(1+1.2*xc)+2.5;
    y=(ysecreto-2.5)+n;
end
```