



Institución
Universitaria
Reacreditada en Alta Calidad

80
Años

Evaluación de Modelos

División del conjunto de datos

Conjuntos de entrenamiento/validación/prueba

La proporción depende en el tamaño del conjunto de datos, pero conjuntos de 80/15/5 usualmente es una buena idea

- El conjunto de entrenamiento es usado para entrenamiento (no es necesario graficar desempeños durante el entrenamiento pero puede ser útil)
- La precisión del conjunto de validación provee un cálculo aproximado del desempeño de generalización (puede tener un sesgo optimista si se diseña la red para un buen rendimiento en el conjunto de validación (“fuga de información”))
- El conjunto de prueba solo debería ser usado una vez para obtener un sesgo estimado del desempeño de generalización

Manejo de los datos en el aprendizaje

Principio general de aprendizaje

Como principio general de aprendizaje, aplicable a todos los modelos:
Sea

$$\mathcal{D} = ((\mathbf{x}^{[1]}, y^{[1]}), (\mathbf{x}^{[2]}, y^{[2]}), \dots, (\mathbf{x}^{[n]}, y^{[n]})) \in (\mathbb{R}^m \times \{0, 1\})^n$$

Modo *On-line*

```
Result:  $\mathbf{w}, b$ 
 $\mathbf{w} := \mathbf{0} \in \mathbb{R}^m, \mathbf{b} := 0;$ 
for  $t = 1, \dots, T$  do
  for  $i = 1, \dots, n$  do
    cálculo de la salida;
    cálculo del error ;
    actualización de parámetros  $\mathbf{w}, b;$ 
  end
end
```

Modo *On-line*

Result: \mathbf{w}, b
 $\mathbf{w} := \mathbf{0} \in \mathbb{R}^m, \mathbf{b} := 0;$
for $t = 1, \dots, T$ **do**
 for $i = 1, \dots, n$ **do**
 cálculo de la salida;
 cálculo del error ;
 actualización de
 parámetros $\mathbf{w}, b;$
 end
end

Modo *On-line II*

Result: \mathbf{w}, b
 $\mathbf{w} := \mathbf{0} \in \mathbb{R}^m, \mathbf{b} := 0;$
for j iteraciones **do**
 Elija un $(\mathbf{x}^{[i]}, y^{[i]}) \in \mathcal{D}$
 aleatorio;
 cálculo de la salida;
 cálculo del error ;
 actualización $\mathbf{w}, b;$
end

Modo *On-line*

Result: \mathbf{w}, b

$\mathbf{w} := \mathbf{0} \in \mathbb{R}^m, \mathbf{b} := 0;$

for $t = 1, \dots, T$ **do**

for $i = 1, \dots, n$ **do**

 cálculo de la salida;

 cálculo del error ;

 actualización de
 parámetros \mathbf{w}, b ;

end

end

Modo *Batch*

Result: \mathbf{w}, b

$\mathbf{w} := \mathbf{0} \in \mathbb{R}^m, \mathbf{b} := 0;$

for $t = 1, \dots, T$ **do**

$\Delta \mathbf{w} := 0, \Delta b := 0;$

for $i = 1, \dots, n$ **do**

 cálculo de la salida;

 cálculo del error ;

 actualización de
 parámetros $\Delta \mathbf{w}, \Delta b$;

end

 actualización de parámetros

\mathbf{w}, b ;

$\mathbf{w} := \mathbf{w} + \Delta \mathbf{w} \quad b := b + \Delta b;$

end

Modo *minibatch*

Combina *On-line* y *Batch*.

$$\mathcal{D} = ((\mathbf{x}^{[1]}, y^{[1]}), (\mathbf{x}^{[2]}, y^{[2]}), \dots, (\mathbf{x}^{[n]}, y^{[n]})) \in (\mathbb{R}^m \times \{0, 1\})^n$$

```
Result:  $\mathbf{w}, b$ 
 $\mathbf{w} := \mathbf{0} \in \mathbb{R}^m, \mathbf{b} := 0;$ 
for  $t = 1, \dots, T$  do
  for  $j = 1, \dots, n/k$  do
     $\Delta \mathbf{w} := \mathbf{0}, \Delta b := 0;$ 
    for  $\{(\mathbf{x}^{[i]}, y^{[i]}), \dots, (\mathbf{x}^{[i+k]}, y^{[i+k]})\} \subset D$  do
      cálculo de la salida;
      cálculo del error ;
      actualización de  $\Delta \mathbf{w}, \Delta b;$ 
    end
    actualización  $\mathbf{w}, b;$ 
     $\mathbf{w} := \mathbf{w} + \Delta \mathbf{w};$ 
     $b := b + \Delta w;$ 
  end
end
```

Medidas de desempeño

Medidas de desempeño a partir de la matriz de confusión

- Exactitud (Accuracy):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Es la proporción de predicciones que el modelo clasificó correctamente.

Medidas de desempeño a partir de la matriz de confusión

- Precisión (Precision):

$$precision = \frac{TP}{TP + FP}$$

Es la proporción de identificaciones positivas que son realmente positivas. O en otras palabras, ¿Qué proporción de clasificaciones positivas fue correcta?

También se conoce como valor predictivo positivo (Positive Predictive Value – PPV)

Medidas de desempeño a partir de la matriz de confusión

- Sensibilidad (Sensitivity):

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

Es la proporción de cuántas muestras positivas fueron clasificadas como positivas.
O en otras palabras, ¿Qué proporción de positivos reales se identificó correctamente?

También se conoce como recall o tasa positiva real (True Positive Rate – TPR)

Medidas de desempeño a partir de la matriz de confusión

- Especificidad (Specificity):

$$specificity = \frac{TN}{TN + FP}$$

Es la proporción de cuántas muestras negativas fueron clasificadas como negativas. O en otras palabras, ¿Qué proporción de negativos reales se identificó correctamente?

También se conoce como selectividad o tasa negativa real (True Negative Rate – TNR)

Medidas de desempeño a partir de la matriz de confusión

- Puntuación F1 (F1-score):

$$F_1score = \frac{2(precision \times recall)}{precision + recall} = \frac{2TP}{2TP + FP + FN}$$

Es la media armónica entre precisión y recall.

Otras métricas: https://en.wikipedia.org/wiki/Confusion_matrix

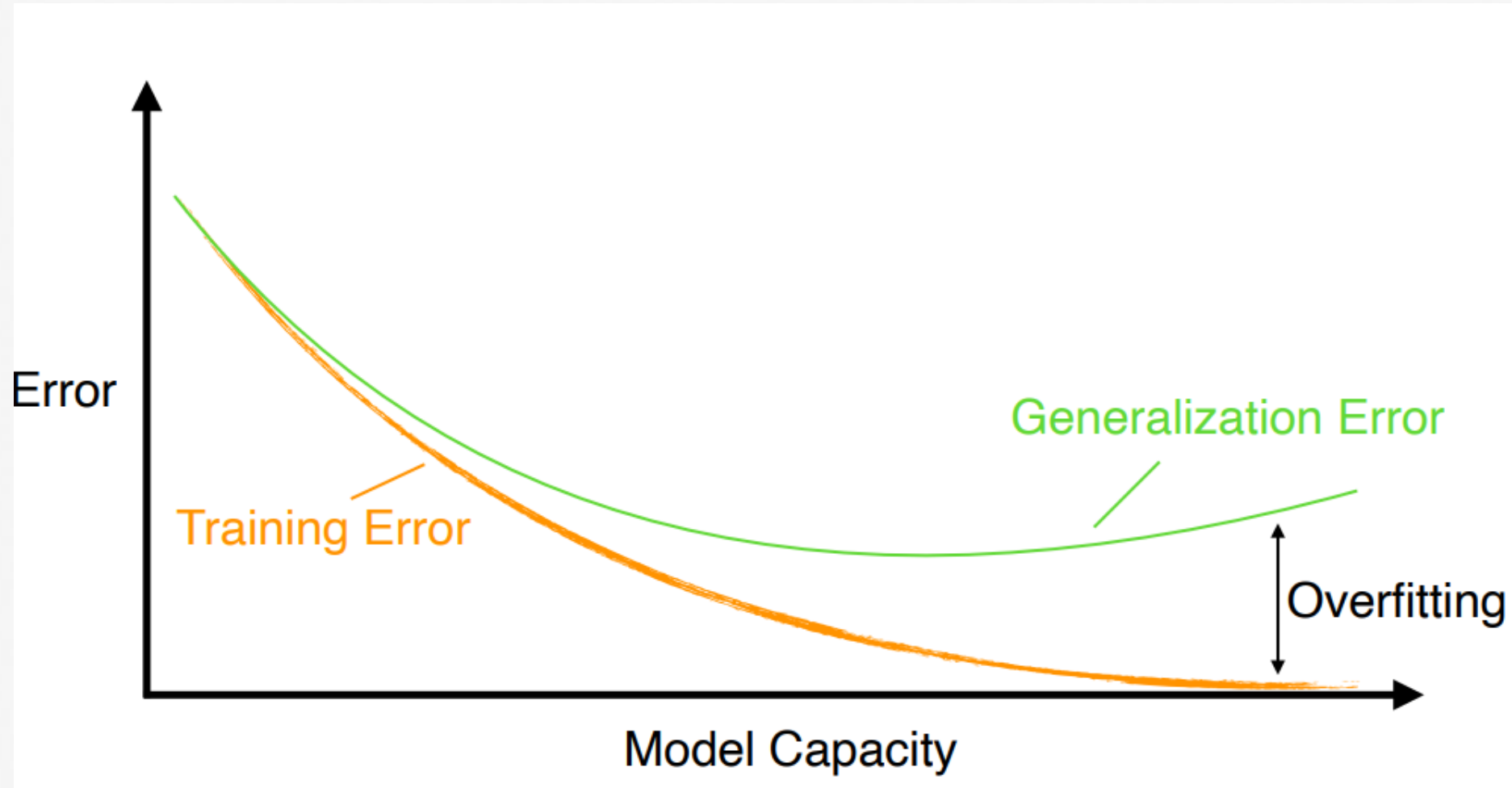
Subentrenamiento y Sobrentrenamiento



El problema con los modelos que son muy simples
y con los modelos que se acoplan “demasiado bien”
a los datos de entrenamiento

Overfitting / Underfitting

Usualmente se usa el error del conjunto de prueba como estimador del error de generalización



Descomposición por sesgo-varianza

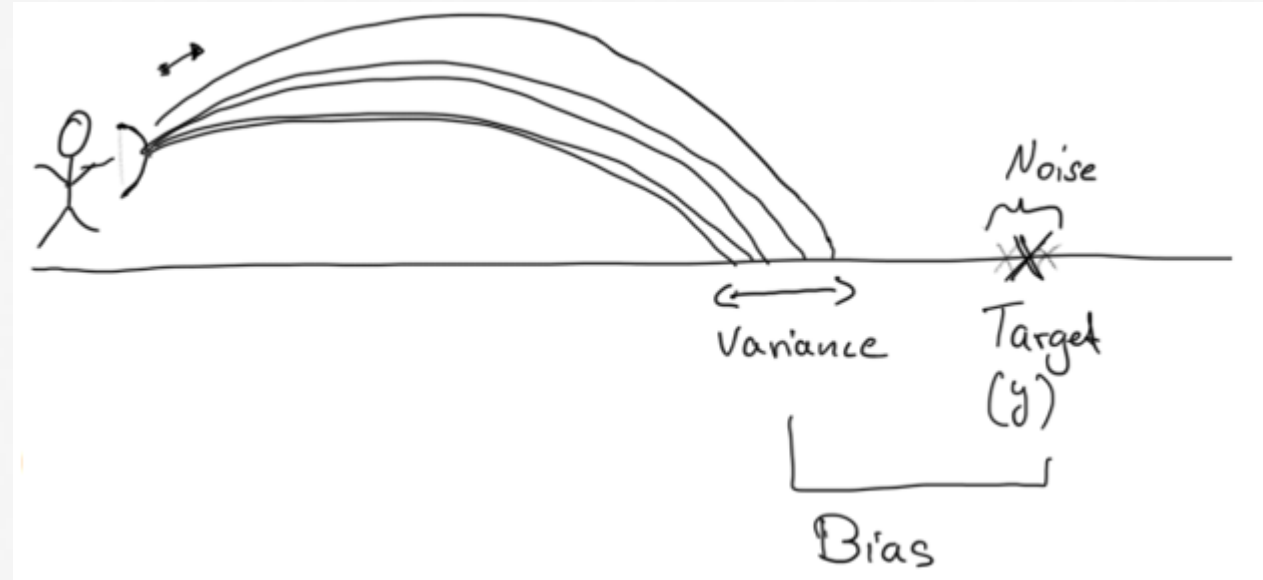
Definición general:

$$\text{Bias}_{\theta}[\hat{\theta}] = E[\hat{\theta}] - \theta$$

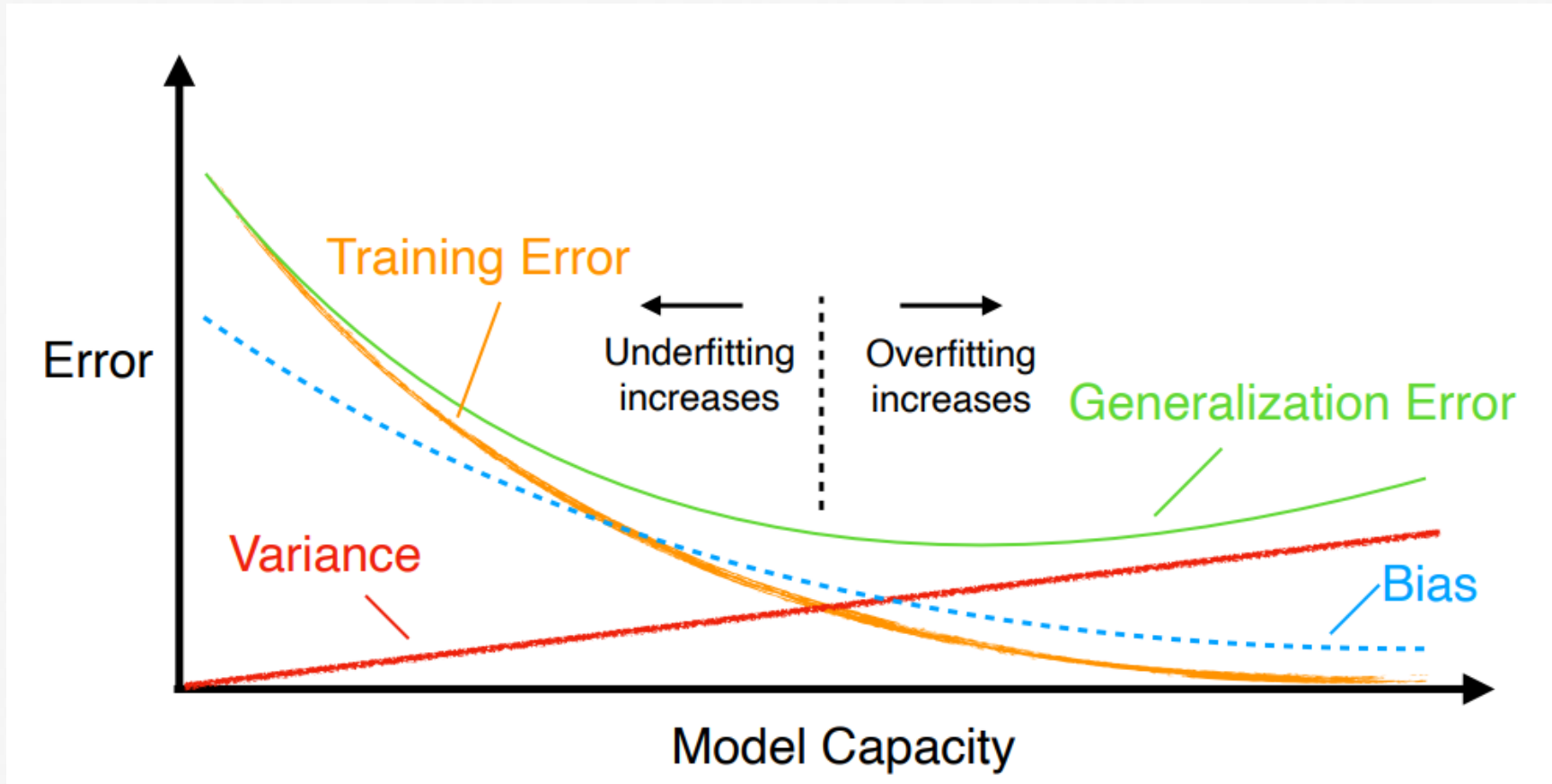
$$\text{Var}_{\theta}[\hat{\theta}] = E[\hat{\theta}^2] - (E[\hat{\theta}])^2$$

$$\text{Var}_{\theta}[\hat{\theta}] = E[(E[\hat{\theta}] - \hat{\theta})^2]$$

Intuición:



Sesgo y Varianza vs Overfitting y Underfitting



Capacidad: concepto abstracto que significa aproximadamente el número de parámetros del modelo multiplicado por la eficiencia con la que se utilizan los parámetros



Objetivo de la regularización: reducir *overfitting*

Generalmente se logra reduciendo la capacidad del modelo o reduciendo la varianza de las predicciones



Institución
Universitaria
Reacreditada en Alta Calidad

80
Años

¡MUCHAS GRACIAS!

