Mecanismos de Votación _

Motivación

Preliminares

- **Objetivo de los ensambles:** mejorar un modelo particular que represente un fenómeno.
- Hasta el momento, los ensambles se fundamentan en la iteración de un modelo específico:
  - **Ensambles Paralelos:** Entrenar múltiples instancias de un modelo y posteriormente promediar.
  - **Ensambles Secuenciales:** Entrenar múltiples instancias de un modelo y corregir de forma iterativa el error predictivo.
- Los ensambles heterogéneos explotan la diversidad de un conjunto de clasificadores débiles para formular un ensamble complejo.

Razones Fundamentales de la Combinación

- Existe una serie de razones para implementar ensambles heterogéneos.
- Éstas se representan en las **razones fundamentales de combinación algorítmica**.
- Diettrich (2000): *Los métodos combinatorios de algoritmos permite fortalecer el proceso de generalización dada tres razones*:
    - Poder Representacional.
    - Poder Estadístico.
    - Poder Computacional.

Razones Fundamentales de la Combinación

$$\mathcal{F} \subset \mathcal{A}_{\mathcal{X}}$$

$$\hat{f} : \mathcal{X} \mapsto \mathcal{A}$$

Razones Fundamentales de la Combinación

$$\mathcal{F} \subset \mathcal{A}_{\mathcal{X}}$$

$$\mathcal{F} = \{\hat{f}_1, \cdots, \hat{f}_{\mathcal{X}}\}$$

$$\hat{f} : \mathcal{X} \mapsto \mathcal{A}$$

Razones Fundamentales de la Combinación

$$\mathcal{F} \subset \mathcal{A}_{\mathcal{X}}$$

$$\mathcal{F} = \{\hat{f}_1, \cdots, \hat{f}_{\mathcal{X}}\} \qquad \hat{f} : \mathcal{X} \mapsto \mathcal{A} \qquad \mathcal{A}_{\mathcal{X}} = \{a_1, \cdots, a_{\mathcal{X}}\}$$

Razones Fundamentales de la Combinación

$$\mathcal{F} \subset \mathcal{A}_{\mathcal{X}}$$

$$\mathcal{F} = \{\hat{f}_1, \cdots, \hat{f}_{\mathcal{X}}\} \qquad \hat{f} : \mathcal{X} \mapsto \mathcal{A} \qquad \mathcal{A}_{\mathcal{X}} = \{a_1, \cdots, a_{\mathcal{X}}\}$$

$$\text{Riesgo}(\hat{f}) = \underset{f \in \mathcal{F}}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ell(f(X_i, y_i))$$

Razones Fundamentales de la Combinación

$$\mathcal{F} \subset \mathcal{A}_{\mathcal{X}}$$

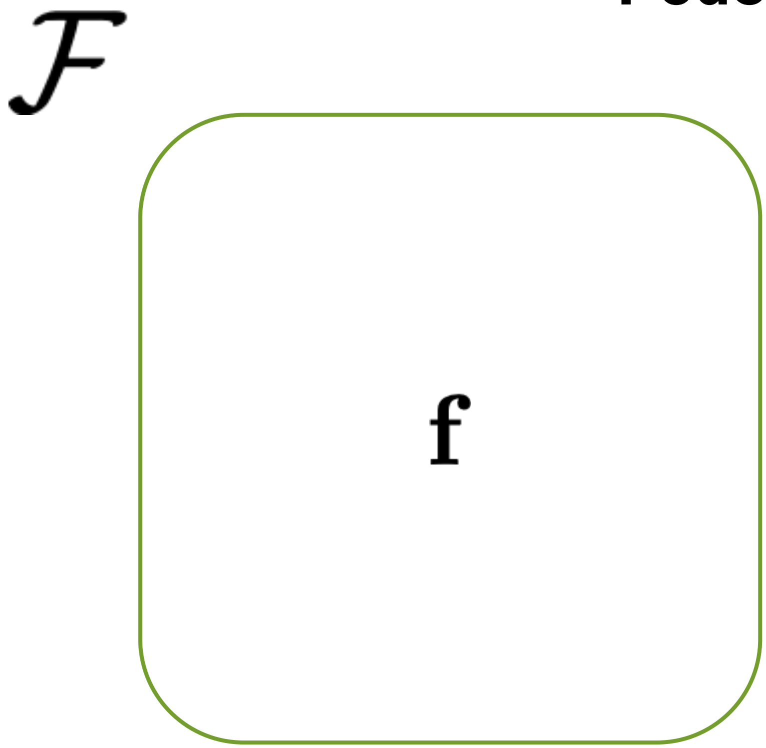
$$\mathcal{F} = \{\hat{f}_1, \cdots, \hat{f}_{\mathcal{X}}\}$$

$$\hat{f} : \mathcal{X} \mapsto \mathcal{A}$$

$$\mathcal{A}_{\mathcal{X}} = \{a_1, \cdots, a_{\mathcal{X}}\}$$

$$\text{Riesgo}(\hat{f}) = \underset{f \in \mathcal{F}}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ell(f(X_i, y_i)$$

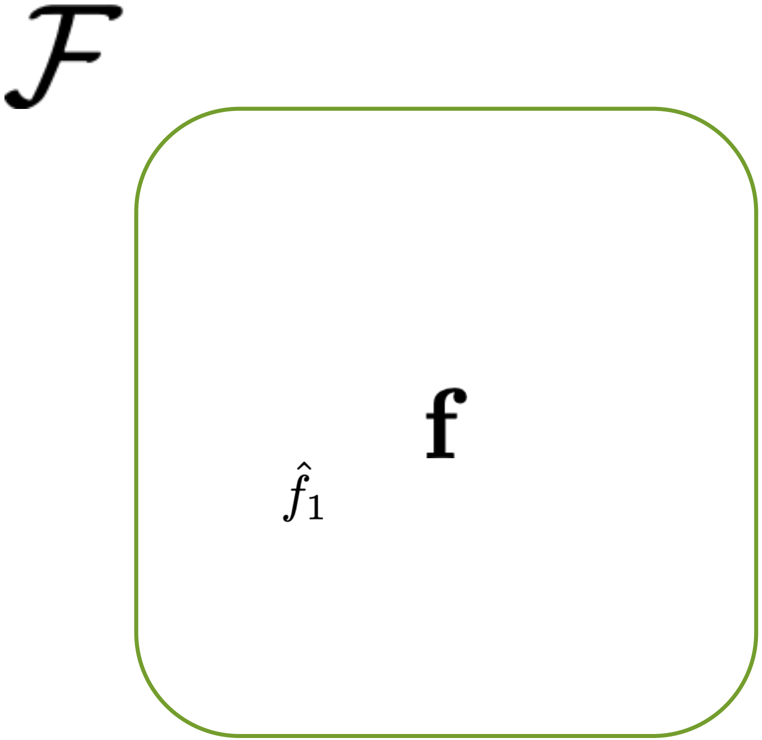
Poder Estadístico



- Puede que el espacio de acción sea muy grande como para que una función singular la capture.

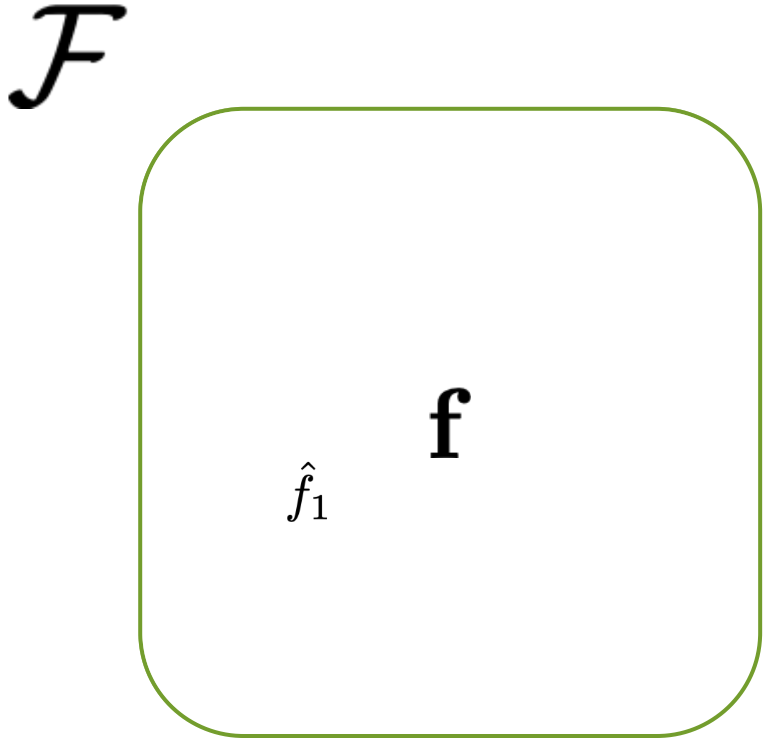
{desafío} latam_

Poder Estadístico



- Puede que el espacio de acción sea muy grande como para que una función singular la capture.
- Múltiples modelos pueden arrojar resultados similares en cuanto a exactitud en el conjunto de entrenamiento.

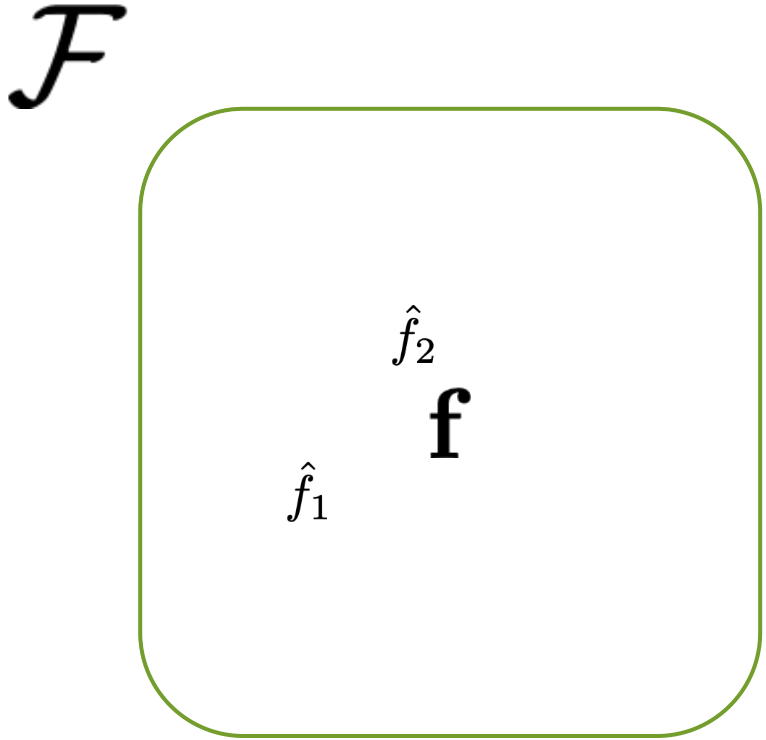
Poder Estadístico



- Puede que el espacio de acción sea muy grande como para que una función singular la capture.
- Múltiples modelos pueden arrojar resultados similares en cuanto a exactitud en el conjunto de entrenamiento.
- Incurrimos en riesgo de falsos positivos del modelo en nuevos datos.

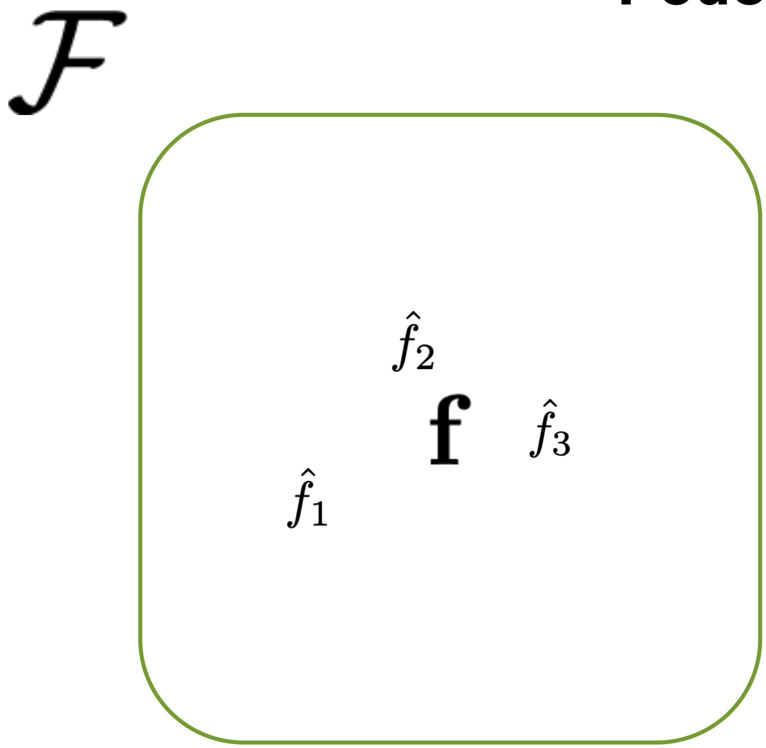
{desafío} latam_

Poder Estadístico



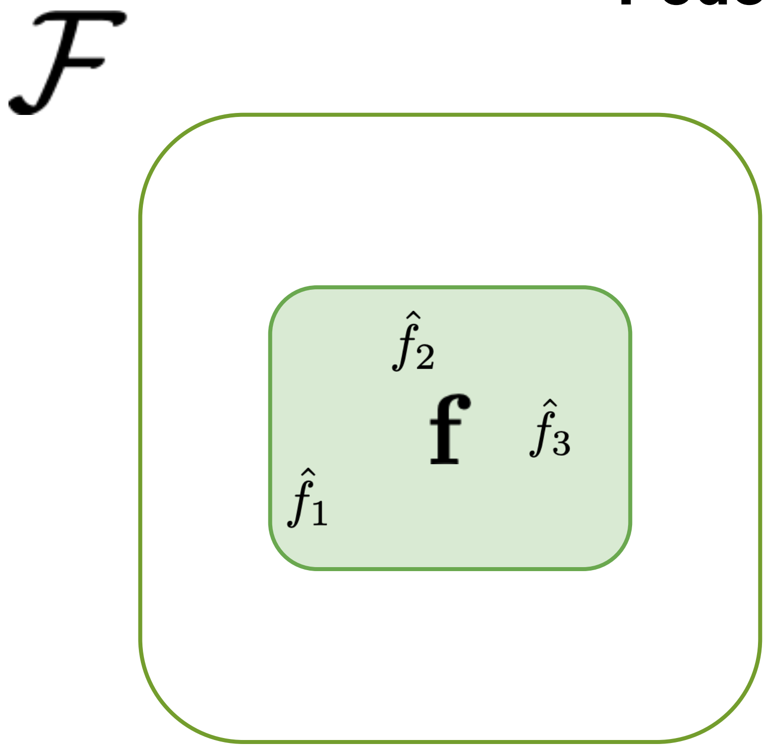
- Puede que el espacio de acción sea muy grande como para que una función singular la capture.
- Múltiples modelos pueden arrojar resultados similares en cuanto a exactitud en el conjunto de entrenamiento.
- Incurrimos en riesgo de falsos positivos del modelo en nuevos datos.
- Mediante ensambles heterogéneo podemos generar una función de búsqueda más compleja

Poder Estadístico



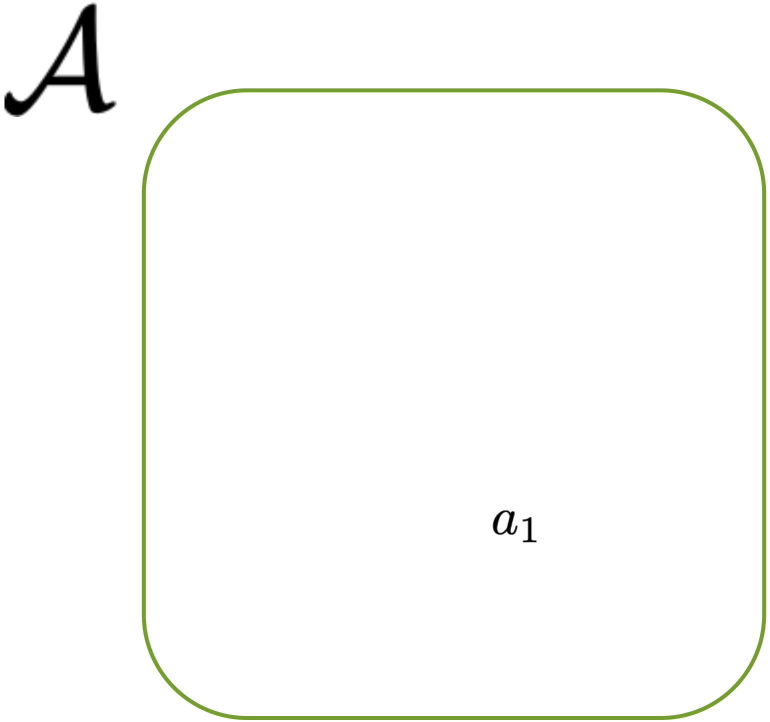
- Puede que el espacio de acción sea muy grande como para que una función singular la capture.
- Múltiples modelos pueden arrojar resultados similares en cuanto a exactitud en el conjunto de entrenamiento.
- Incurrimos en riesgo de falsos positivos del modelo en nuevos datos.
- Mediante ensambles heterogéneo podemos generar una función de búsqueda más compleja

Poder Estadístico



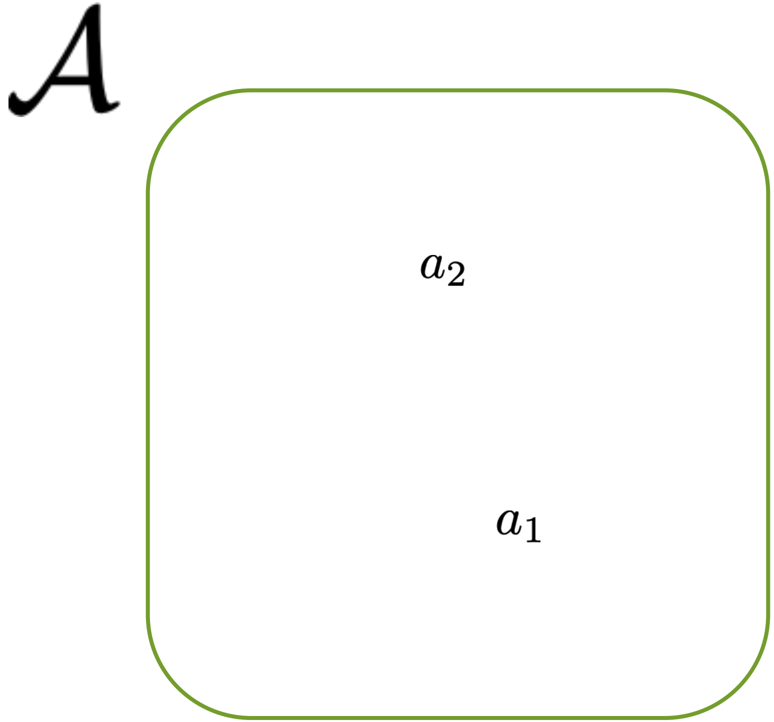
- Puede que el espacio de acción sea muy grande como para que una función singular la capture.
- Múltiples modelos pueden arrojar resultados similares en cuanto a exactitud en el conjunto de entrenamiento.
- Incurrimos en riesgo de falsos positivos del modelo en nuevos datos.
- Mediante ensambles heterogéneo podemos generar una función de búsqueda más compleja

Poder Representacional



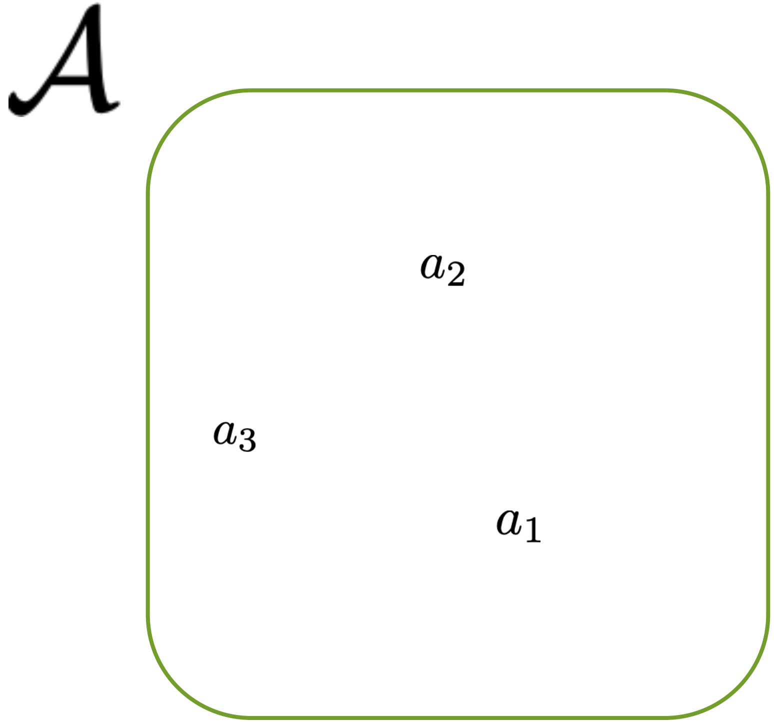
- En muchas tareas de Machine Learning tenemos una hipótesis de trabajo que es desconocida.

Poder Representacional



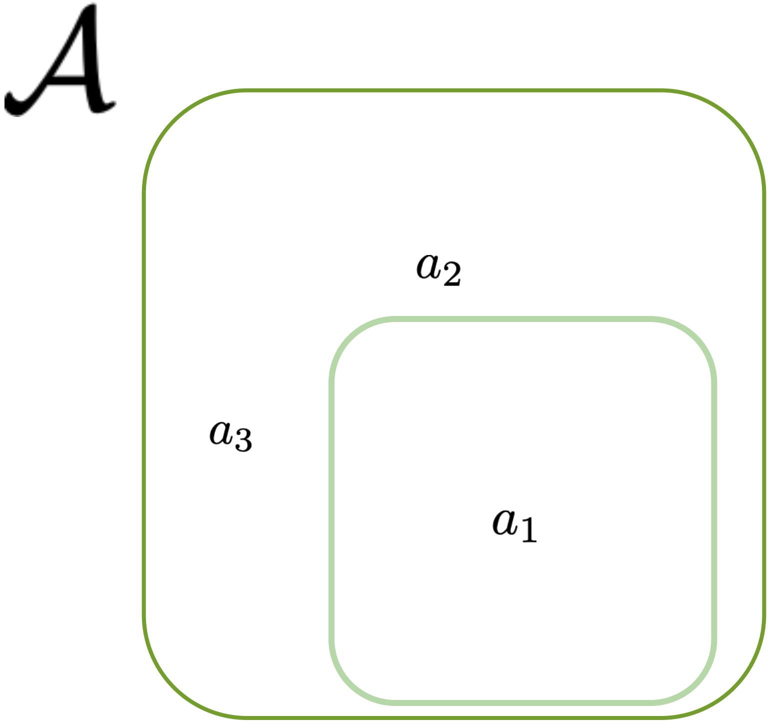
- En muchas tareas de Machine Learning tenemos una hipótesis de trabajo que es desconocida.
- Esta no puede ser representada completamente por una hipótesis de acción específica.

Poder Representacional



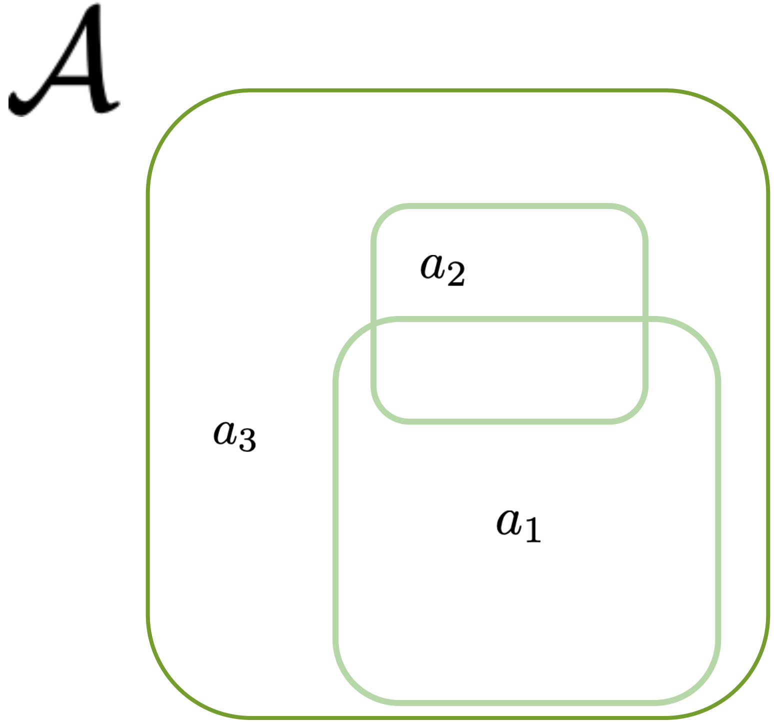
- En muchas tareas de Machine Learning tenemos una hipótesis de trabajo que es desconocida.
- Esta no puede ser representada completamente por una hipótesis de acción específica.

Poder Representacional



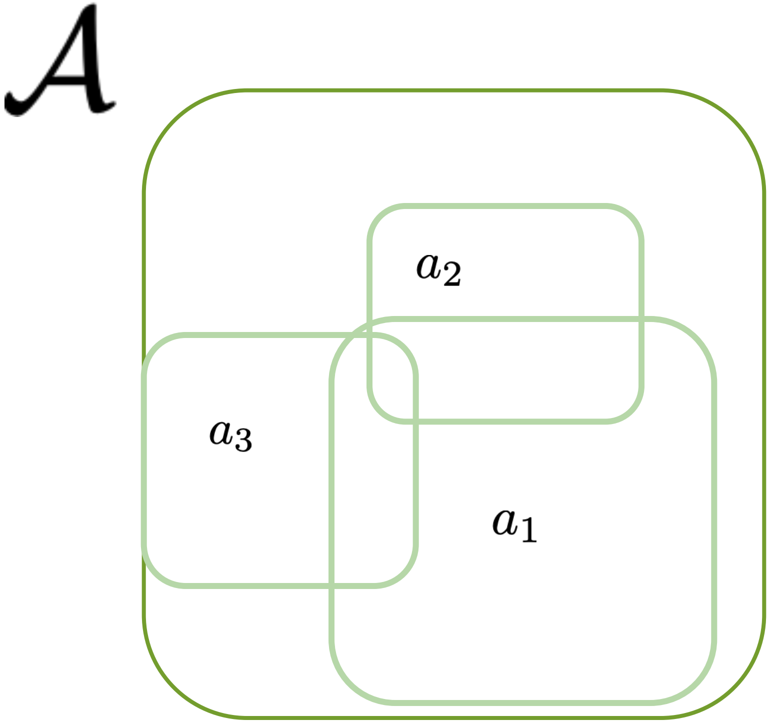
- En muchas tareas de Machine Learning tenemos una hipótesis de trabajo que es desconocida.
- Esta no puede ser representada completamente por una hipótesis de acción específica.
- Mediante la combinación de algoritmos, logramos la expansión del espacio de funciones representables.

Poder Representacional



- En muchas tareas de Machine Learning tenemos una hipótesis de trabajo que es desconocida.
- Esta no puede ser representada completamente por una hipótesis de acción específica.
- Mediante la combinación de algoritmos, logramos la expansión del espacio de funciones representables.

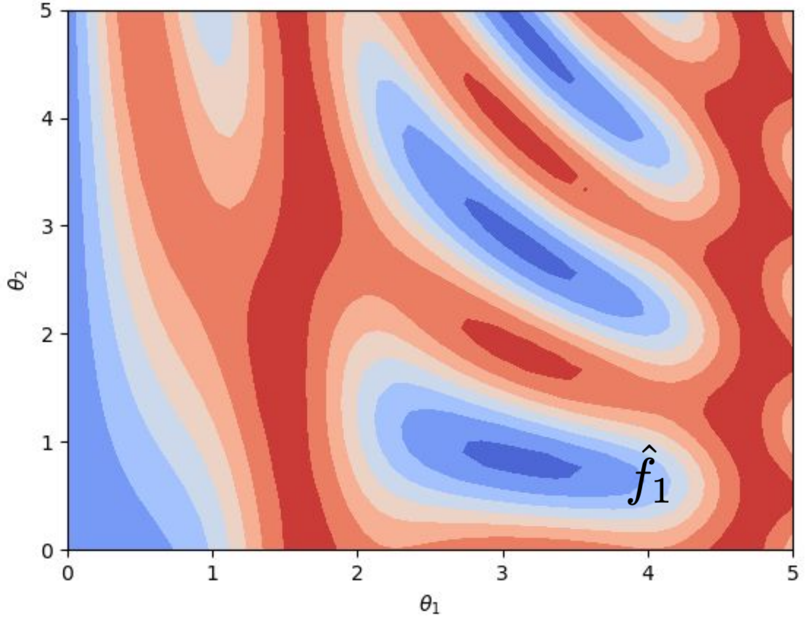
Poder Representacional



- En muchas tareas de Machine Learning tenemos una hipótesis de trabajo que es desconocida.
- Esta no puede ser representada completamente por una hipótesis de acción específica.
- Mediante la combinación de algoritmos, logramos la expansión del espacio de funciones representables.

Poder Computacional

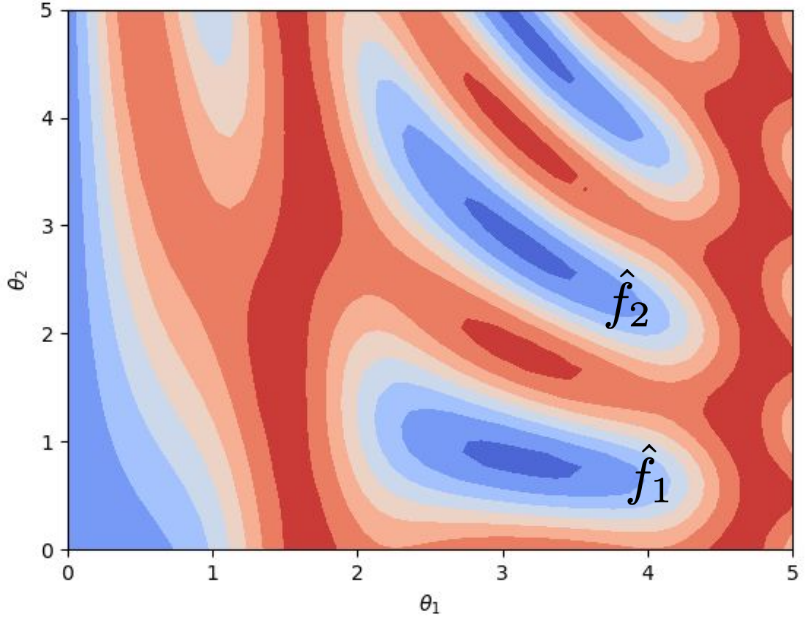
$$\mathbb{E}[\ell(\theta)]$$



- Algunos algoritmos de instancia única pueden sufrir de ineficiencia en el proceso de optimización.

Poder Computacional

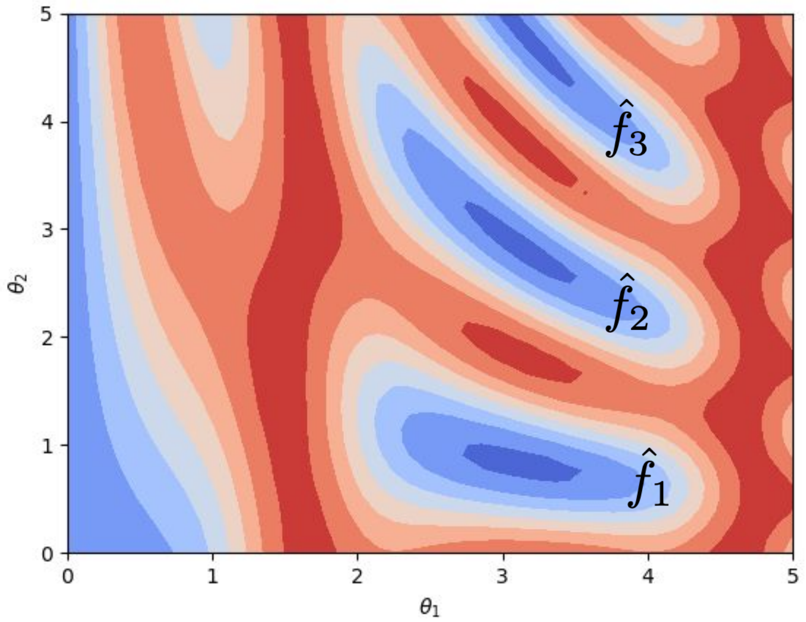
$$\mathbb{E}[\ell(\theta)]$$



- Algunos algoritmos de instancia única pueden sufrir de ineficiencia en el proceso de optimización.
- Al ensamblar múltiples modelos, podemos evaluar el comportamiento de la superficie de optimización en múltiples puntos.

Poder Computacional

$$\mathbb{E}[\ell(\theta)]$$



- Algunos algoritmos de instancia única pueden sufrir de ineficiencia en el proceso de optimización.
- Al ensamblar múltiples modelos, podemos evaluar el comportamiento de la superficie de optimización en múltiples puntos.
- De esta manera, el proceso permite evitar el problema de estacionarse en un mínimo local en nuestra superficie de la función de pérdida.

Mecanismos de Voto

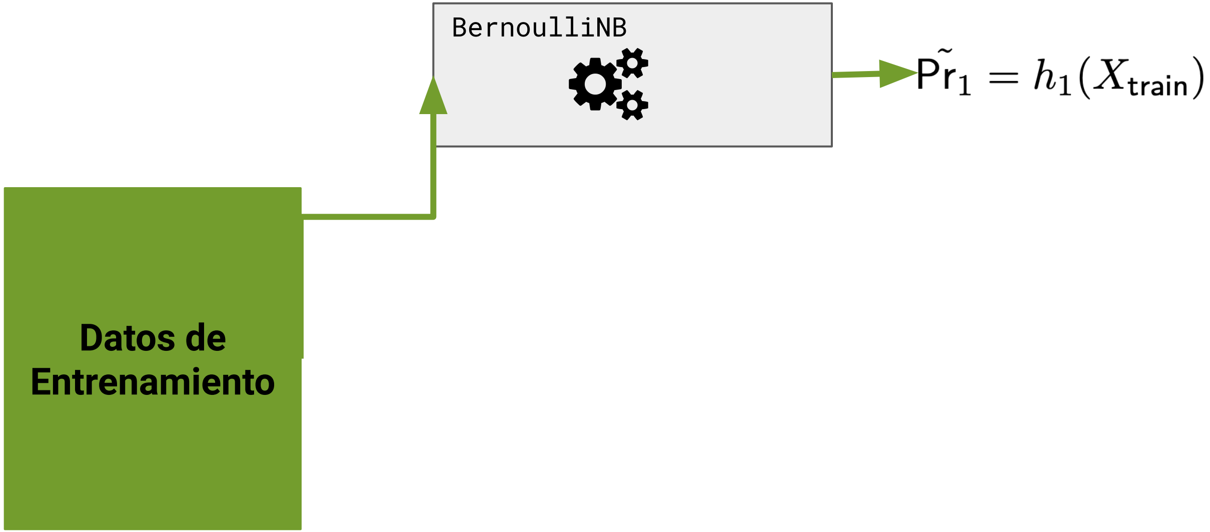
Rudimentos

- A grandes rasgos, el proceso de votación se puede reducir a los siguientes pasos:
  - Entrenar cada modelo por sí solo y preservar las predicciones
  - Promediar las predicciones siguiendo algún esquema.
- La promediación de predicciones responde a la siguiente estructura:
  - Tipo del vector objetivo.
  - Mecanismo de decisión.
- Para un problema de **regresión**, el mecanismo es un promedio.
- Para un problema de **clasificación**, el mecanismo puede ser mayoría simple o relativa.
- Ambos enfoques se pueden ponderar mediante la relevancia del modelo, la cual es un criterio de investigador.

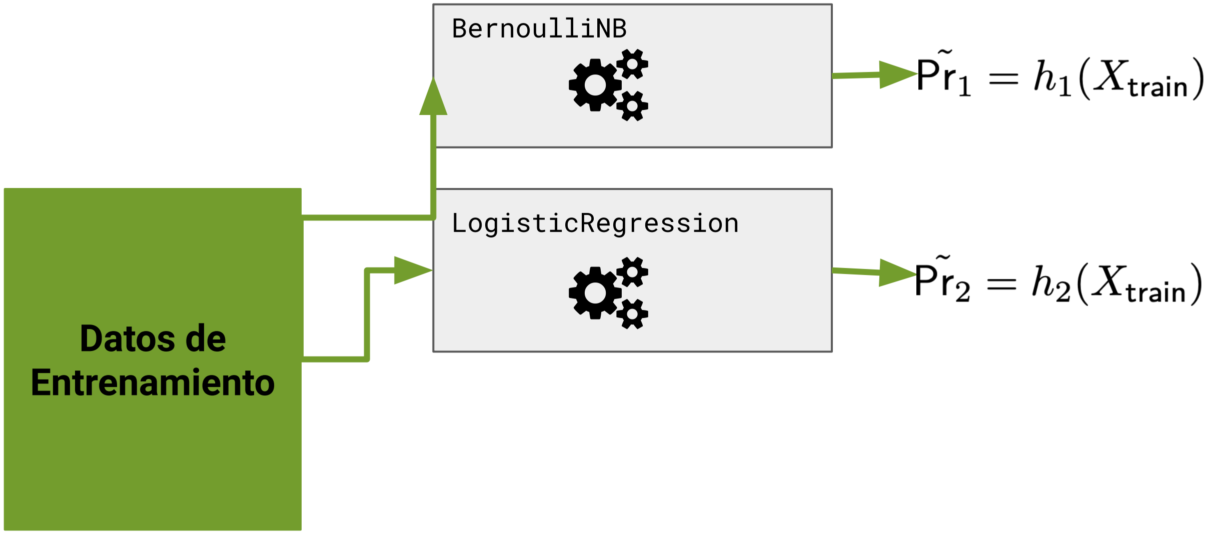
Mecanismo de votación

Datos de Entrenamiento

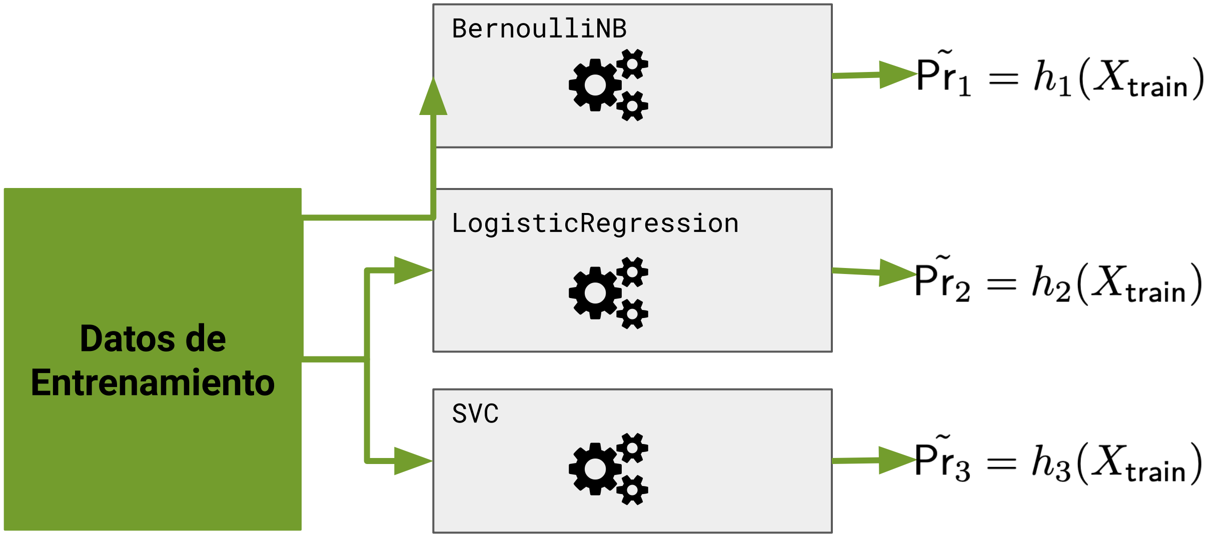
Mecanismo de votación



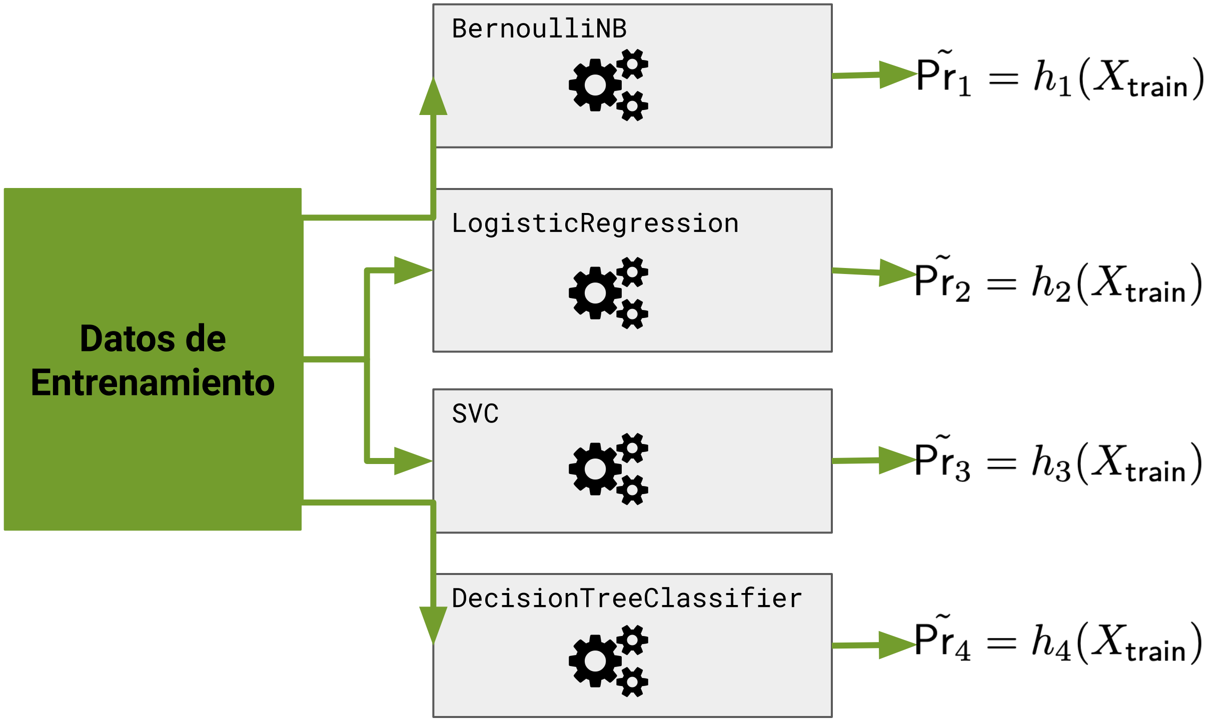
Mecanismo de votación



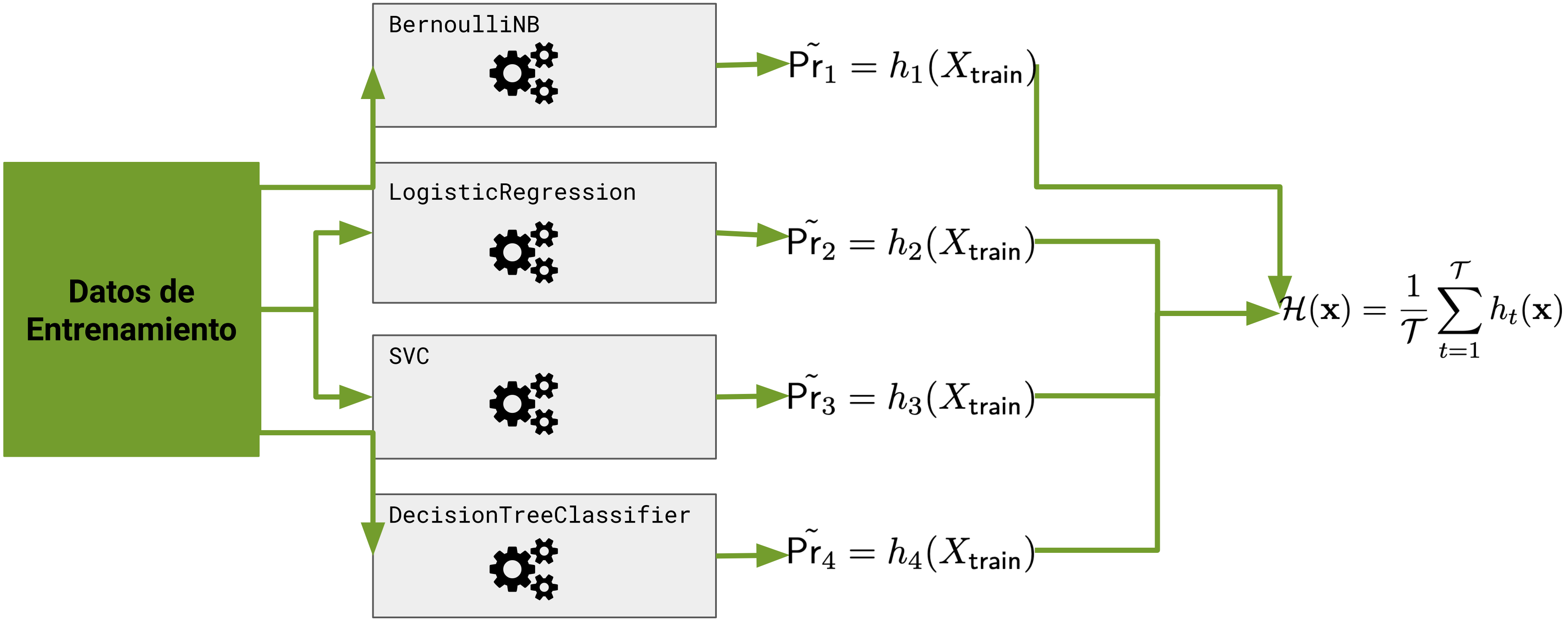
Mecanismo de votación



Mecanismo de votación



Mecanismo de votación



Comité con vector objetivo contínuo

$$\mathcal{H}(\mathbf{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} h_t(\mathbf{x})$$

- La estrategia más común es implementar un promedio simple en las predicciones realizadas.

Comité con vector objetivo contínuo

$$\mathcal{H}(\mathbf{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} h_t(\mathbf{x})$$

$$h_t(\mathbf{x}) = f(\mathbf{x}) + \varepsilon_t(\mathbf{x}), \quad \forall t \in \mathcal{T}$$

- La estrategia más común es implementar un promedio simple en las predicciones realizadas.
- Parte de las virtudes de implementar un ensamble heterogéneo es el hecho de desagregar el valor final en dos partes.

Comité con vector objetivo contínuo

$$\mathcal{H}(\mathbf{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} h_t(\mathbf{x})$$

$$h_t(\mathbf{x}) = f(\mathbf{x}) + \varepsilon_t(\mathbf{x}), \quad \forall t \in \mathcal{T}$$

- La estrategia más común es implementar un promedio simple en las predicciones realizadas.
- Parte de las virtudes de implementar un ensamble heterogéneo es el hecho de desagregar el valor final en dos partes.
- Tendremos una parte fija que representará el “valor verdadero” y un error contextual a cada miembro del ensamble.

{desafío} latam_

Comité con vector objetivo discreto

El primer punto a considerar es si vamos a promediar clases u etiquetas.

$$\mathcal{H}(\mathbf{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} h_t(\mathbf{x})$$

Comité con vector objetivo discreto

El primer punto a considerar es si vamos a promediar clases u etiquetas.

Si promediamos etiquetas, el resultado provendrá de la probabilidad específica de ocurrencia para una clase.

$$\mathcal{H}(\mathbf{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} h_t(\mathbf{x})$$

Etiqueta: $h_t^c(\mathbf{x}) \in [0, 1] \to \mathrm{Pr}(c|\mathbf{x})$

{desafío} latam_

Comité con vector objetivo discreto

El primer punto a considerar es si vamos a promediar clases u etiquetas.

Si promediamos etiquetas, el resultado provendrá de la probabilidad específica de ocurrencia para una clase.

Si promediamos clase, el resultado provendrá de algún criterio de decisión arbitrario.

$$\mathcal{H}(\mathbf{x}) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} h_t(\mathbf{x})$$

Etiqueta: $h_t^c(\mathbf{x}) \in [0,1] \rightarrow \mathrm{Pr}(c|\mathbf{x})$

Clase: $h_t^c(\mathbf{x}) \in \{0,1\}$

{desafío} latam_

Comité con vector objetivo discreto

$$\mathcal{H}(\mathbf{x}) = \begin{cases} C & \text{si } \sum_{t=1}^{T} h_t^j(\mathbf{x}) > \frac{1}{2}\sum_{k=1}^{l}\sum_{t=1}^{T} h_t^k(\mathbf{x}) \\ \text{rechazo} & \text{de lo contrario} \end{cases}$$

Cuando nuestro vector objetivo es discreto y nuestra estrategia es de maximización de clase con un 50%+1, podemos optar por la siguiente heurística de mayoría simple.

Comité con vector objetivo discreto

$$\mathcal{H}(\mathbf{x}) = \underset{c}{\text{argmax}} \sum_{t=1}^{\mathcal{T}} h_t^c(\mathbf{x})$$

Cuando nuestro vector objetivo es discreto y nuestra estrategia es de maximización de la mayor clase, podemos optar por la siguiente heurística de mayoría relativa..

Relevancia del clasificador en el comité

$$\mathsf{Pr}_{mv} = \sum_{k=[t/2+1]}^{\mathcal{T}} \binom{\mathcal{T}}{k} \pi^k (1-\pi)^{\mathcal{T}-k}$$

Podemos extraer la relevancia (cantidad de predicciones correctas en una decisión de ensamble) de un clasificador mediante la siguiente expresión.

Relevancia del clasificador en el comité

$$\mathsf{Pr}_{mv} = \sum_{k=[t/2+1]}^{\mathcal{T}} \binom{\mathcal{T}}{k} \pi^{k}(1-\pi)^{\mathcal{T}-k}$$

Mediante el coeficiente binomial entre la cantidad de miembros del comité (T) y clases (k), podemos ajustar la probabilidad de éxito/fracaso de un modelo específico

Relevancia del clasificador en el comité

$$\text{Pr}_{mv} = \sum_{k=[t/2+1]}^{\mathcal{T}} \binom{\mathcal{T}}{k} \pi^{k} (1-\pi)^{\mathcal{T}-k}$$

Estimamos las predicciones realizadas para la clase **específica**.

Relevancia del clasificador en el comité

$$\Pr_{mv} = \sum_{k=[t/2+1]}^{\mathcal{T}} \binom{\mathcal{T}}{k} \pi^{k} (1-\pi)^{\mathcal{T}-k}$$

Estimamos las predicciones realizadas para la clase **remanente**.

Relevancia del clasificador en el comité

$$\text{Pr}_{mv} = \sum_{k=[t/2+1]}^{\mathcal{T}} \binom{\mathcal{T}}{k} \pi^{k} (1-\pi)^{\mathcal{T}-k}$$

Sumamos a nivel de clases estimables.

Esquema de Ponderación

{desafío} latam_

Mecanismo de ponderación

$$\mathcal{H}(\mathbf{x}) = \operatorname*{argmax}_{c} \sum_{t=1}^{\mathcal{T}} w_i h_t^c(\mathbf{x})$$

- Cuando nuestro modelo se guíe por mayoría absoluta (50% + 1), podemos asignar relevancia específica a cada ensamble.

Mecanismo de ponderación

$$\mathcal{H}(\mathbf{x}) = \underset{c}{\text{argmax}} \sum_{t=1}^{\mathcal{T}} w_i h_t^c(\mathbf{x})$$

$$\sum_{i=1}^{N} w_i \equiv 1$$

- Cuando nuestro modelo se guíe por mayoría absoluta (50% + 1), podemos asignar relevancia específica a cada ensamble.
- Mediante la ponderación, definimos ex ante el comportamiento de cada clasificador.

{desafío} latam_

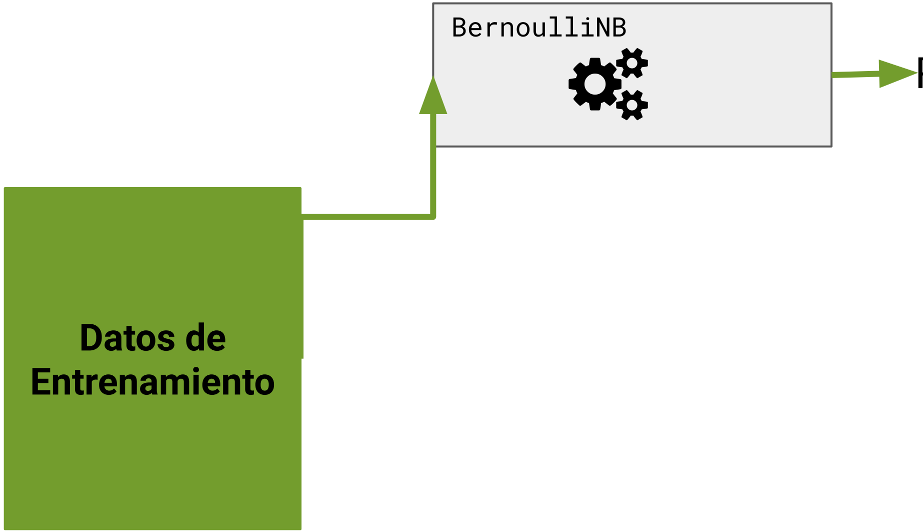
Mecanismo de ponderación

$$\mathcal{H}(\mathbf{x}) = \underset{c}{\text{argmax}} \sum_{t=1}^{\mathcal{T}} w_i h_t^c(\mathbf{x})$$

$$\sum_{i=1}^{N} w_i \equiv 1$$

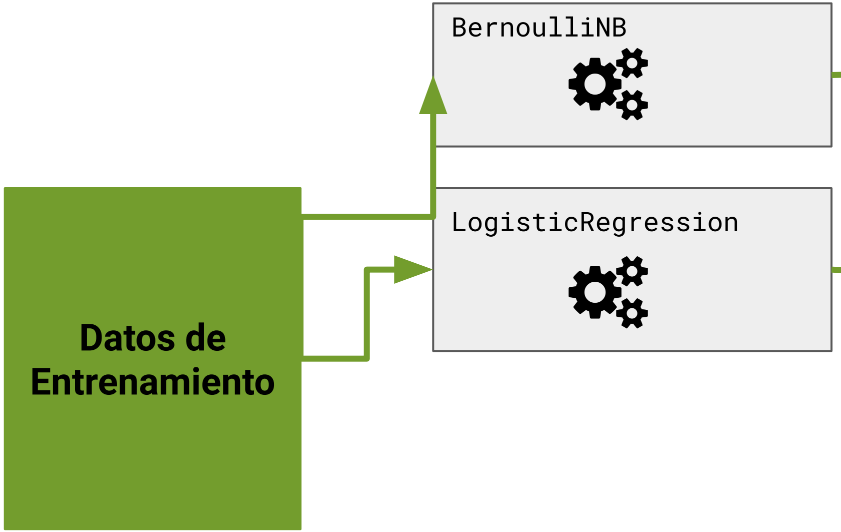
- Cuando nuestro modelo se guíe por mayoría absoluta (50% + 1), podemos asignar relevancia específica a cada ensamble.
- Mediante la ponderación, definimos ex ante el comportamiento de cada clasificador.
- Los ponderadores usualmente deben reflejar los axiomas de Kolmogorov.

Votación en un esquema ponderado



$$\tilde{\text{Pr}}_1 = h_1(X_{\text{Train}}) \times w_1$$

Votación en un esquema ponderado

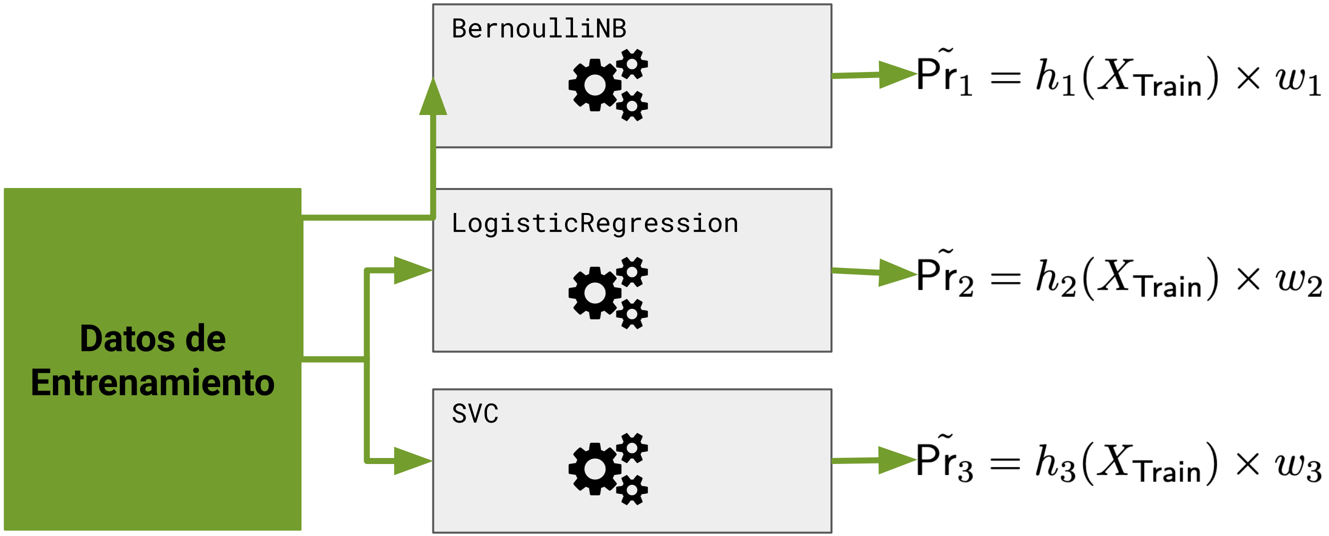


$$\tilde{\Pr}_1 = h_1(X_{\text{Train}}) \times w_1$$

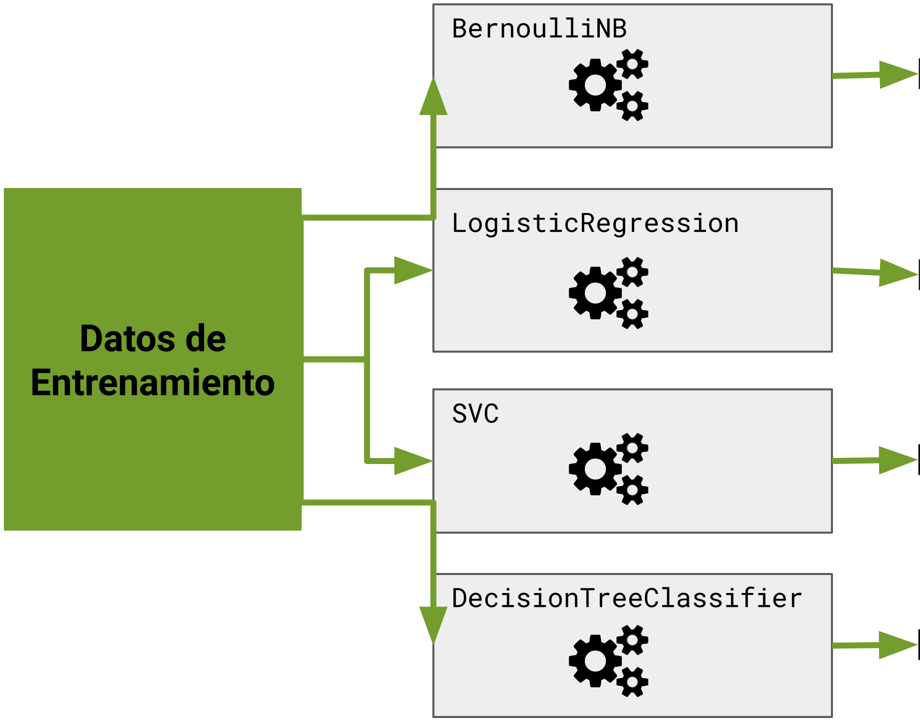
$$\tilde{\Pr}_2 = h_2(X_{\text{Train}}) \times w_2$$

{desafío} latam_

Votación en un esquema ponderado



Votación en un esquema ponderado



$$\tilde{\Pr}_1 = h_1(X_{\text{Train}}) \times w_1$$

$$\tilde{\Pr}_2 = h_2(X_{\text{Train}}) \times w_2$$

$$\tilde{\Pr}_3 = h_3(X_{\text{Train}}) \times w_3$$

$$\tilde{\Pr}_4 = h_4(X_{\text{Train}}) \times w_4$$

{desafío} latam_

Votación en un esquema ponderado

{desafío} latam_

{desafío}
latam_

Academia de
talentos digitales

www.desafiolatam.com