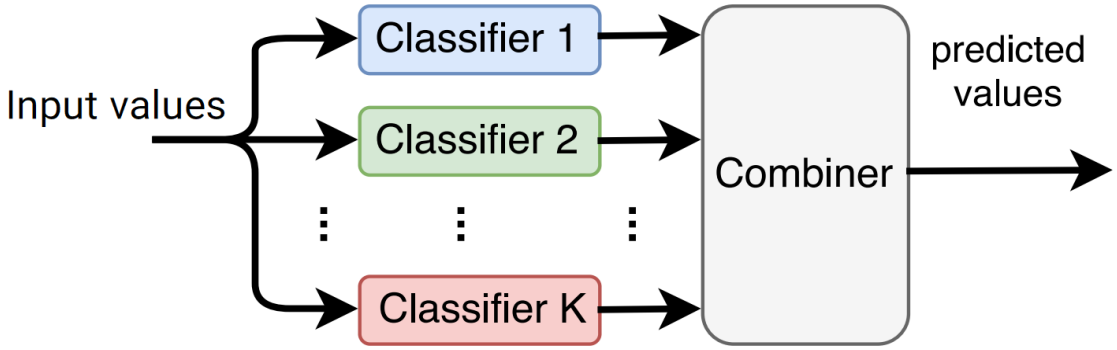
یادگیری گروهی

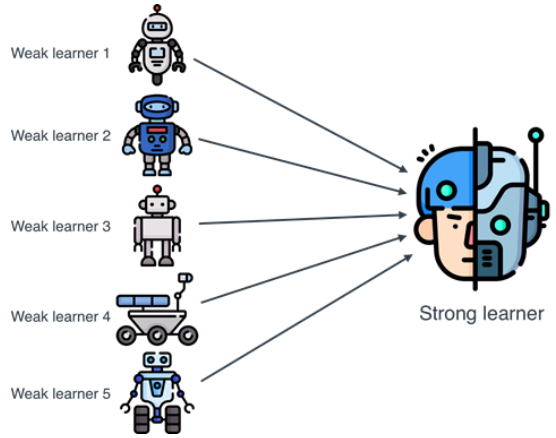
Ensemble Learning

dayche.com . گروه دایکه

یادگیری گروهی

- یادگیری گروهی یک اصطلاح کلی برای متدهایی است که برای تصمیم‌گیری از ترکیب چندین مدل استفاده می‌کنند (معمولا در وظایف یادگیری ماشین با نظارت).





با ترکیب چندین مدل، اشتباهات هر مدل توسط سایر مدل‌ها جبران می‌شود و در نتیجه، عملکرد کلی پیش‌بینی یادگیری گروهی بهتر از یک مدل واحد خواهد بود.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

یادگیری گروهی

یادگیری گروهی را می‌توان تفسیر یادگیری ماشین از خرد جمعی بیان کرد.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

یادگیری گروهی

ماجرای پلیموث

در سال 1906 در یک نمایشگاه در پلیموث انگلستان، 800 نفر در مسابقه تخمین وزن گاو ذبح شده شرکت کردند.

فیلسوف و ریاضی‌دان انگلیسی فرانسیس گالتون متوجه شد که مقدار میانه نظرات (1207 پوند) این 800 نفر با اختلاف کمتر از 1% از وزن واقعی گاو (1198 پوند) قرار دارد.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

یادگیری گروهی

Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?

Manuel Fernández-Delgado MANUEL.FERNANDEZ.DELGADO@USC.ES
Eva Cernadas EVA.CERNADAS@USC.ES
Senén Barro SENEN.BARRO@USC.ES
CITIUS: Centro de Investigación en Tecnoloxías da Información da USC
University of Santiago de Compostela
Campus Vida, 15872, Santiago de Compostela, Spain
Dinani Amorim DINANIAMORIM@GMAIL.COM
Departamento de Tecnologia e Ciências Sociais- DTCS
Universidade do Estado da Bahia
Av. Edgard Chastinet S/N - São Geraldo - Juazeiro-BA, CEP: 48.305-680, Brasil

Editor: Russ Greiner

Abstract

We evaluate **179 classifiers** arising from **17 families** (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest-neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods), implemented in Weka, R (with and without the caret package), C and Matlab, including all the relevant classifiers available today. We use **121 data sets**, which represent **the whole UCI** data base (excluding the large-scale problems) and other own real problems, in order to achieve significant conclusions about the classifier behavior, not dependent on the data set collection. **The classifiers most likely to be the bests are the random forest** (RF) versions, the best of which (implemented in R and accessed via caret) achieves 94.1% of the maximum accuracy overcoming 90% in the 84.3% of the data sets. However, the difference is not statistically significant with the second best, the SVM with Gaussian kernel implemented in C using LibSVM, which achieves 92.3% of the maximum accuracy. A few models are clearly better than the remaining ones: random forest, SVM with Gaussian and polynomial kernels, extreme learning machine with Gaussian kernel, C5.0 and avNNet (a committee of multi-layer perceptrons implemented in R with the caret package). The random forest is clearly the best family of classifiers (3 out of 5 bests classifiers are RF), followed by SVM (4 classifiers in the top-10), neural networks and boosting ensembles (5 and 3 members in the top-20, respectively).

Keywords: classification, UCI data base, random forest, support vector machine, neural networks, decision trees, ensembles, rule-based classifiers, discriminant analysis, Bayesian classifiers, generalized linear models, partial least squares and principal component regression, multiple adaptive regression splines, nearest-neighbors, logistic and multinomial regression

توانایی بالای الگوریتم‌های مبتنی بر یادگیری گروهی

امروزه الگوریتم‌های مبتنی‌بر یادگیری گروهی در بسیاری از وظایف یادگیری ماشین قدرتمند عمل می‌کنند و انتخاب نهایی بسیاری از محققین این حوزه است.

فرناندز دلگادو و همکارانش (2014) در مقاله‌ای به بررسی و مقایسه 179 الگوریتم از 17 خانواده یادگیری ماشین روی 121 دیتاست پرداختند. در این مقاله الگوریتم‌های مبتنی‌بر یادگیری گروهی بهترین نتایج را داشتند.

یادگیری گروهی

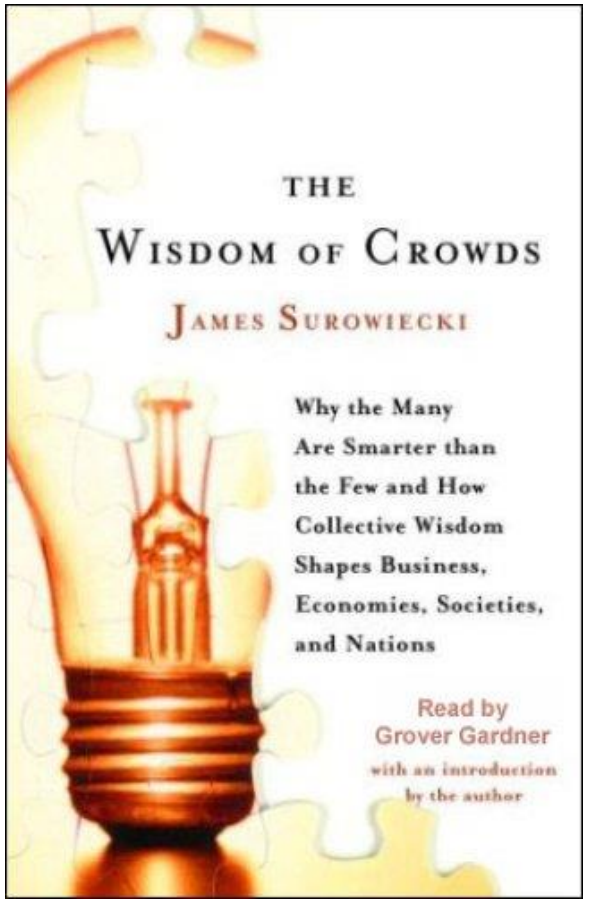
پایه های یادگیری گروهی

یادگیری گروهی به تولید و ترکیب چندین مدل جهت حل یک وظیفه خاص اشاره دارد.

این تفسیر شهودی از متدولوژی یادگیری گروهی ریشه در طبعیت انسان و تمایل وی برای جمع‌آوری نظرات مختلف و وزن‌دهی و ترکیب آن‌ها برای تصمیم‌گیری‌های پیچیده دارد.

وزن‌دهی و تجمیع چندین نظر شخصی متفاوت بهتر از انتخاب یک نظر شخصی است.

یادگیری گروهی



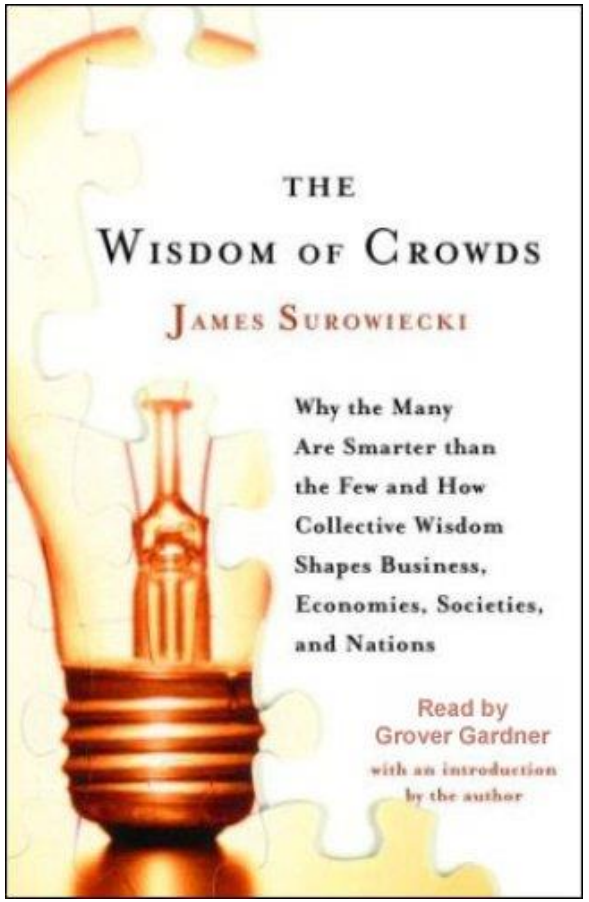
این نظریه دو محدودیت بزرگ داشت:

- فرض استقلال رای‌دهندگان
- باینری بودن مسئله تصمیم‌گیری

جیمز مایکل سرویسکی در کتاب "خرد جمعی: چرا اکثریّت باهوش‌تر از اقلیّت هستند" توضیح می‌دهد که چگونه می‌توان اطلاعات چندین منبع را ترکیب کرد طوری‌که نتیجه آن اغلب از یک منبع بهتر باشد.

باید توجه داشت که به‌طور طبیعی همیشه خروجی جمع خردمندانه نیست!

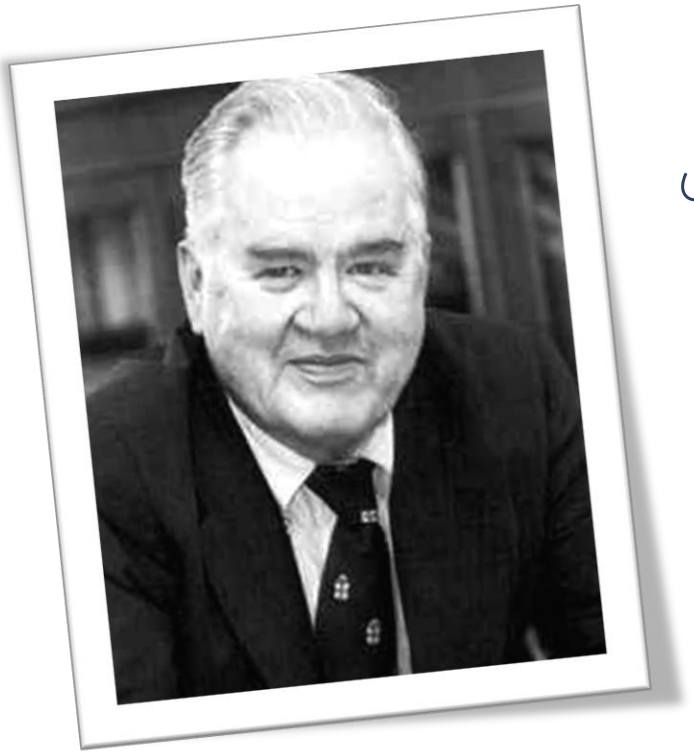
یادگیری گروهی



از منظر سرویسکی خرد جمعی زمانی از یک تصمیم‌گیری انفرادی احتمالا پیشی می‌گیرید که ضوابط زیر ارضا شده باشد:

- **استقلال**: نظر هر فرد مستقل از نظرات دیگران باشد
- **تمرکز زدایی**: فرد قادر به تصمیم‌گیری بر اساس اطلاعات محلی خود است
- **تنوع نظرات**: وجود افراد با تخصص‌ها و بینش‌های متفاوت در نظرات
- **تجمیع**: وجود مکانیزم‌هایی برای قضاوت خصوصی از مجموعه‌ای تصمیم شخصی

یادگیری گروهی



پیدایش یادگیری گروهی

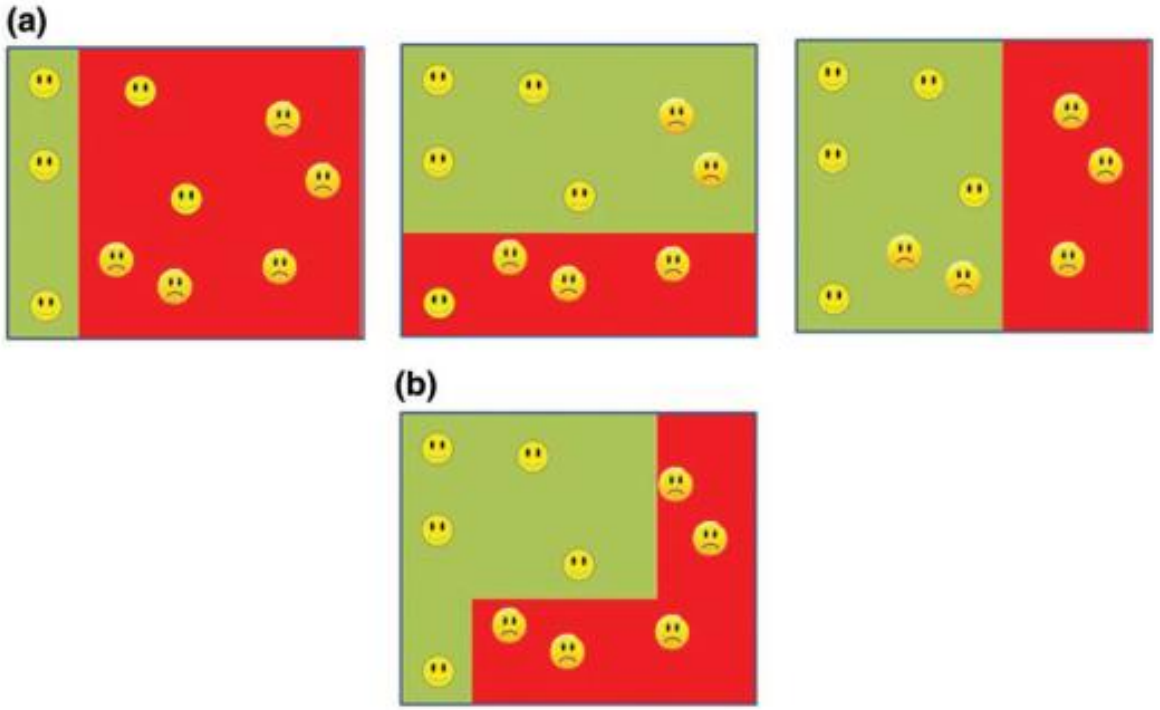
در دهه 1970 میلادی جان توکی ریاضی‌دان بزرگ امریکایی و خالق بسیاری از مفاهیم حوزه کامپیوتر اولین مدل یادگیری گروهی را با ترکیب در مدل خطی ارائه داد.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

یادگیری گروهی

چرا یادگیری گروهی خوب عمل می کند؟!

- جلوگیری از بیش‌برازش
- فواید محاسباتی
- بازنمایی

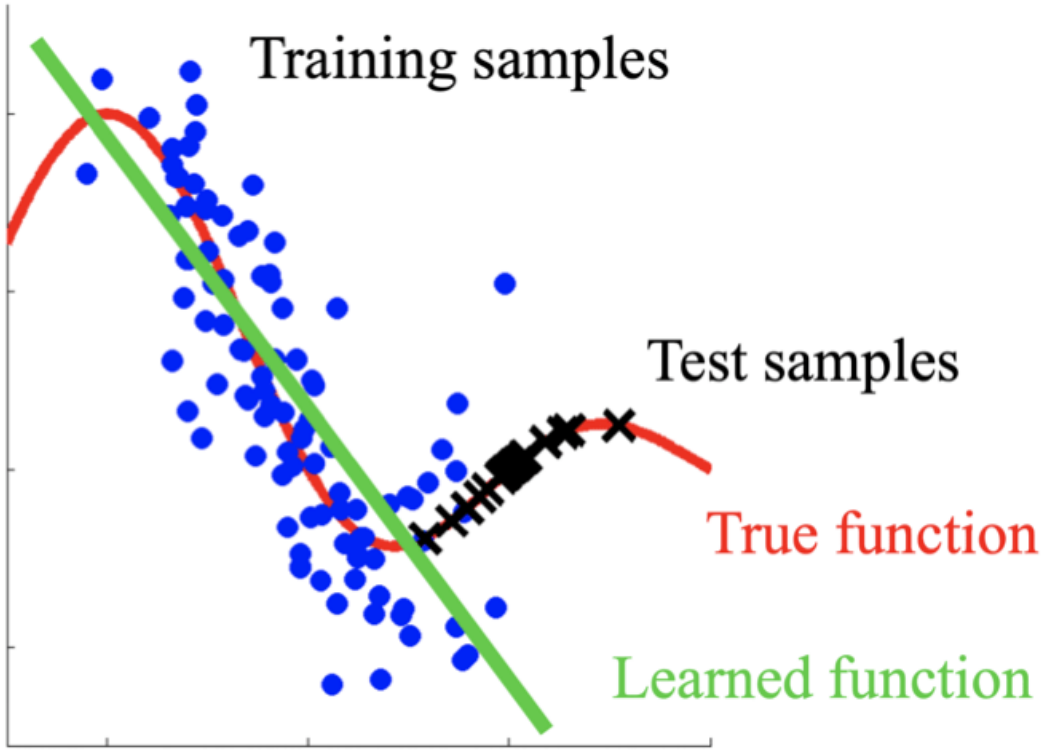


تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

یادگیری گروهی

یادگیری ترکیبی ممکن است جهت حل بعضی مشکلات در یادگیری ماشین استفاده شود:

- کلاس‌های نامتعادل (Class Imbalance)
- رانش مفهوم (Concept Drift)
- طلسم ابعاد (Curse of Dimensionality)



تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

ساخت یک مدل یادگیر گروهی

فرض کنید دیتاست ما دارای n نمونه و هر نمونه دارای m ویژگی است:

$$D = \{(x_i, y_i)\}\ (|D| = n, x_i \in R^m, y_i \in R)$$

مدل یادگیر ترکیبی φ از ترکیب k مدل یادگیرنده $\{f_1, f_2, \dots, f_k\}$ و اعمال تابع تجمیع G جهت پیش‌بینی یک خروجی طبق فرمول زیر تعریف شده است:

$$\hat{y}_i = \varphi(x_i) = \mathrm{G}(f_1, f_2, \dots, f_k)$$

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

آموزش مدل

چندین راه برای آموزش یک مدل ترکیبی و دست‌یابی به یک عملکرد خوب وجود دارد. با این حال بعضی مشخصه‌های اصلی در زمان ایجاد یک مدل ترکیبی باید در نظر گرفته شود:

- **تنوع**: عملکرد خوب مدل‌های ترکیبی به طور عمده به خاطر استفاده از سوءگیری‌های استقرایی مختلف است. بنابراین، مدل‌ها در مدل ترکیبی می‌بایست به‌طور موثری متنوع باشد تا بتوان عملکرد نهایی خوبی بدست آورد.
- **عملکرد پیش‌بینی‌کننده**: عملکرد پیش‌بینی مدل‌های درون مدل ترکیبی باید تا حد امکان قدرتمند باشد

آموزش مدل

چندین رویکرد برای دست‌یابی به تنوع در مدل‌ها وجود دارد:

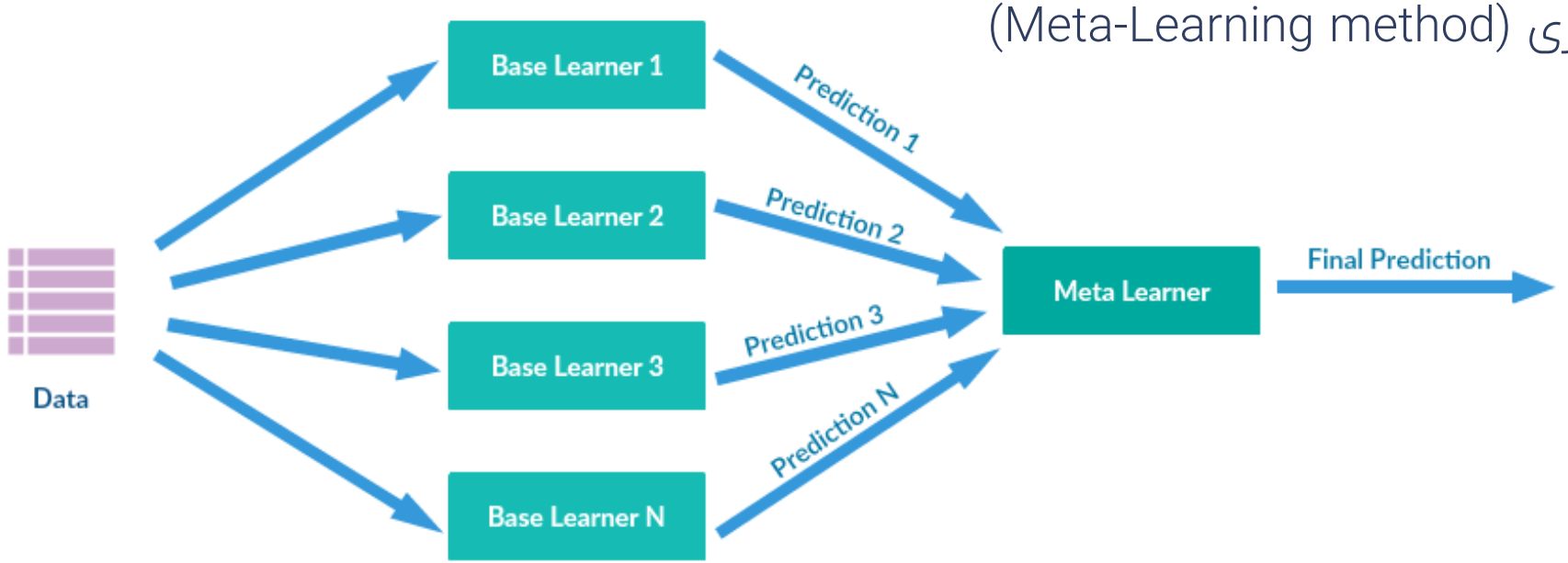
- دست‌کاری ورودی (Input manipulation)
- الگوریتم یادگیری دست‌کاری‌شده (Manipulated learning algorithm)
- پارتیشن‌کردن (Partitioning)
- دست‌کاری خروجی (Output manipulation)
- ترکیب چند استراتژی (Ensemble hybridization)

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

تجمیع خروجی ها

چندین رویکرد برای تجمیع خروجی‌های مدل‌ها در روش ترکیبی وجود دارد:

- روش وزن‌دهی (Weighting methods)
- روش فرا-یادگیری (Meta-Learning method)



تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

روش‌های یادگیری گروهی را را می‌توان در دو چارچوب اصلی تقسیم کرد:

- **چارچوب وابسته (Dependent framework)**

خروجی هر مدل روی ساخت مدل بعد تاثیر می‌گذارد. به‌بیان دیگر، دانش ایجاد شده در مرحله قبل راهنمای یادگیری در مرحله بعد است.

- **چارچوب مستقل (Independent framework)**

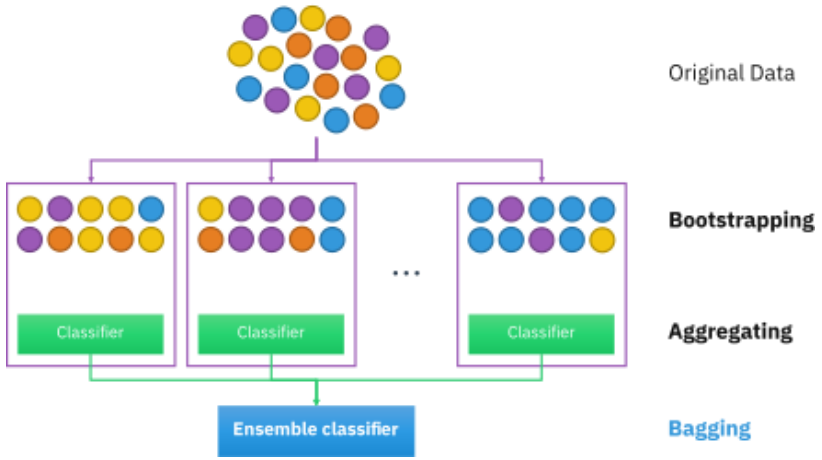
هر مدل مستقل از مدل‌های دیگر آموزش داده می‌شود.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

روش Bagging (Bootstrap Aggregating)

یکی از ساده‌ترین رویکردها برای ایجاد یک مدل مستقل گروهی است که در آن هر یادگیرنده با نمونه‌های متفاوتی آموزش داده می‌شود.



Bagging Training

Input: I (a base inducer), T (the number of iterations), S (the original training set), μ (the sample size).

1: $t \leftarrow 1$
2: **repeat**
3: $S_t \leftarrow$ a sample of μ instances from S with replacement.
4: Construct classifier M_t using I, with S_t as the training set
5: $t \leftarrow t+1$
6: **until** $t > T$

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

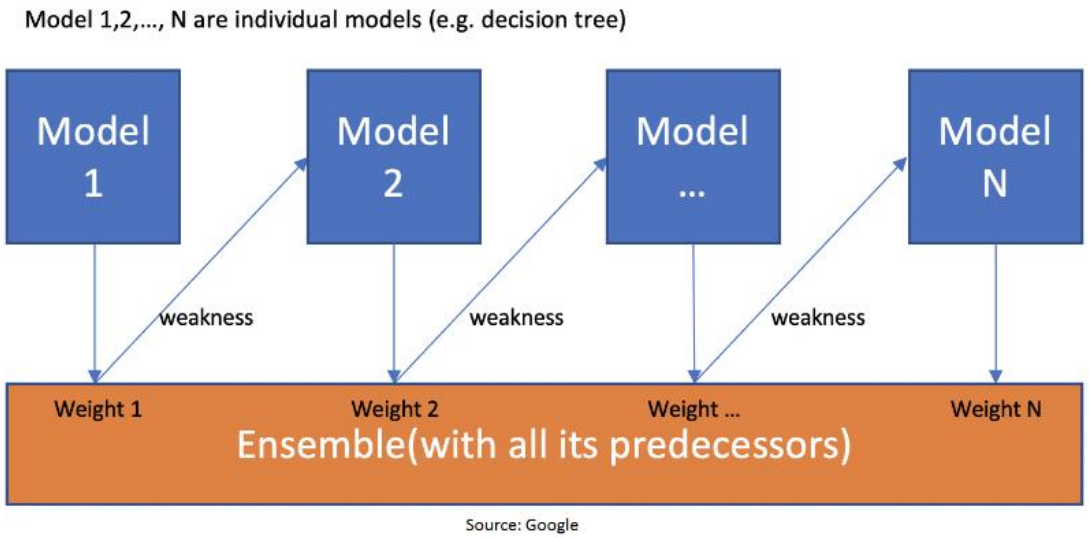
روش AdaBoost

- معروف‌ترین الگوریتم وابسته برای ساخت مدل گروهی است.
- ایده اصلی آن تمرکز روی نمونه‌هایی است که یادگیرنده قبلی به اشتباه آن‌ها را دسته‌بندی کرده است.
- میزان تمرکز روی نمونه‌ها با اختصاص وزن‌دهی به هر نمونهٔ آموزشی انجام می‌شود.
- در مرحله اول، وزن تمام نمونه‌ها ثابت در نظر گرفته می‌شود.
- در هر مرحله وزن نمونه‌هایی که به اشتباه دسته‌بندی شده‌اند افزایش و وزن نمونه‌هایی که درست دسته‌بندی شده‌اند کاهش می‌یابد.
- بعلاوه، به هر یادگیرنده بر اساس عملکرد کلی آن وزنی اختصاص داده می‌شود.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

روش AdaBoost



Source: Google

AdaBoost Training

Input: I (a weak inducer), T (the number of iterations), and S (a training set).

Output: $M_t, \alpha_t; t = 1, \ldots, T$

1: $t \leftarrow 1$
2: $D_1(i) \leftarrow 1/m; i = 1, \ldots, m$
3: **repeat**
4: Build Classifier M_t using I and distribution D_t
5: $\varepsilon_t \leftarrow \sum_{i:M_t(x_i) \neq y_i} D_t(i)$
6: **if** $\varepsilon_t > 0.5$ **then**
7: $T \leftarrow t - 1$
8: exit Loop.
9: **end if**
10: $\alpha_t \leftarrow \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$
11: $D_{t+1}(i) = D_t(i) \cdot e^{-\alpha_t y_t M_t(x_i)}$
12: Normalize D_{t+1} to be a proper distribution.
13: $t \leftarrow t + 1$
14: **until** $t > T$

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

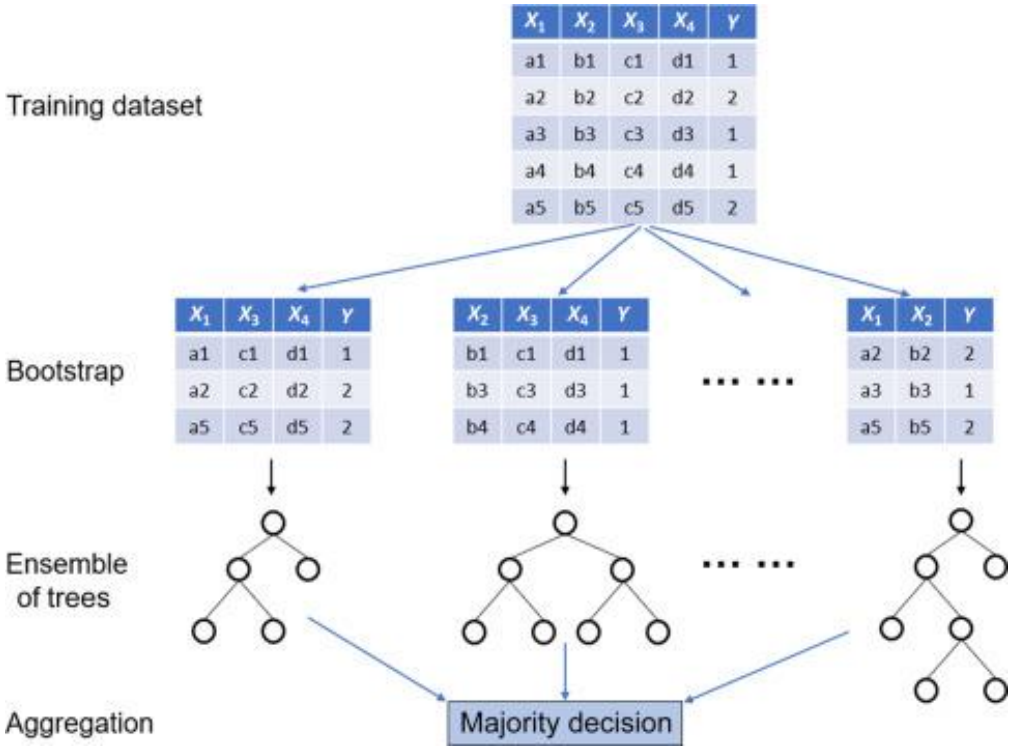
روش های یادگیری گروهی

روش Random Forest

- معروف‌ترین الگوریتم مستقل برای ساخت مدل گروهی است.
- علّت افزایش روزافزون محبوبیت این روش:
  - سادگی
  - عملکرد قوی
  - پیاده‌سازی و آموزش ساده آن
- این روش از تعداد زیادی درخت تصمیم حرص نشده استفاده می‌کند.
- جهت رشد یک درخت تصادفی که کماکان دقت مناسبی داشته باشد از دو فرآیند تصادفی، عدم قطعیت را وارد درخت‌های تصمیم می‌کنند:
  - هر درخت را با مجموعه نمونه‌های متفاوتی آموزش می‌دهند
  - در هر مرحله به جای انتخاب بهترین نود در درخت (موثرترین ویژگی)، به صورت تصادفی یک مجموعه از ویژگی‌ها، انتخاب شده و از بین آن‌ها بهترین نود انتخاب می‌شود (می‌تواند متفاوت هم باشد!).

روش های یادگیری گروهی

روش Random Forest



Input: IDT (a decision tree inducer), T (the number of iterations), S (the training set), μ (the subsample size), and N (the number of attributes used in each node).

Output: $M_t; t = 1, \ldots, T$

1: $t \leftarrow 1$
2: **repeat**
3: $S_t \leftarrow$ Sample μ instances from S with replacement.
4: Build classifier M_t using $IDT(N)$ on S_t
5: $t++$
6: **until** $t > T$

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

روش Gradient Boosting Machines

- در این روش، آموزش هر یادگیرنده وابسته به یادگیرنده‌ای است که در مرحله قبل آموزش دیده است
- تفاوت اصلی بین GBM و دیگر تکنیک‌ها این است که در GBM بهینه‌سازی روی فضای تابع اعمال شده است.
- این روش شامل یک رویه یادگیری است که در آن هدف ساخت بهترین یادگیرنده‌هایی است به‌نحوی که آن‌ها بیشینه ارتباط را با منفی گرادیان تابع هزینه‌ای که روی کل مدل گروهی اختصاص یافته را داشته باشند.
- به بیان دقیق‌تر، در GBM یک توالی از درخت‌های رگرسیونی محاسبه می‌شود که هر درخت متوالی شبه‌خطای درخت‌های قبلی را با توجه به یک تابع خطای مشتق‌پذیر پیش‌بینی کند.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

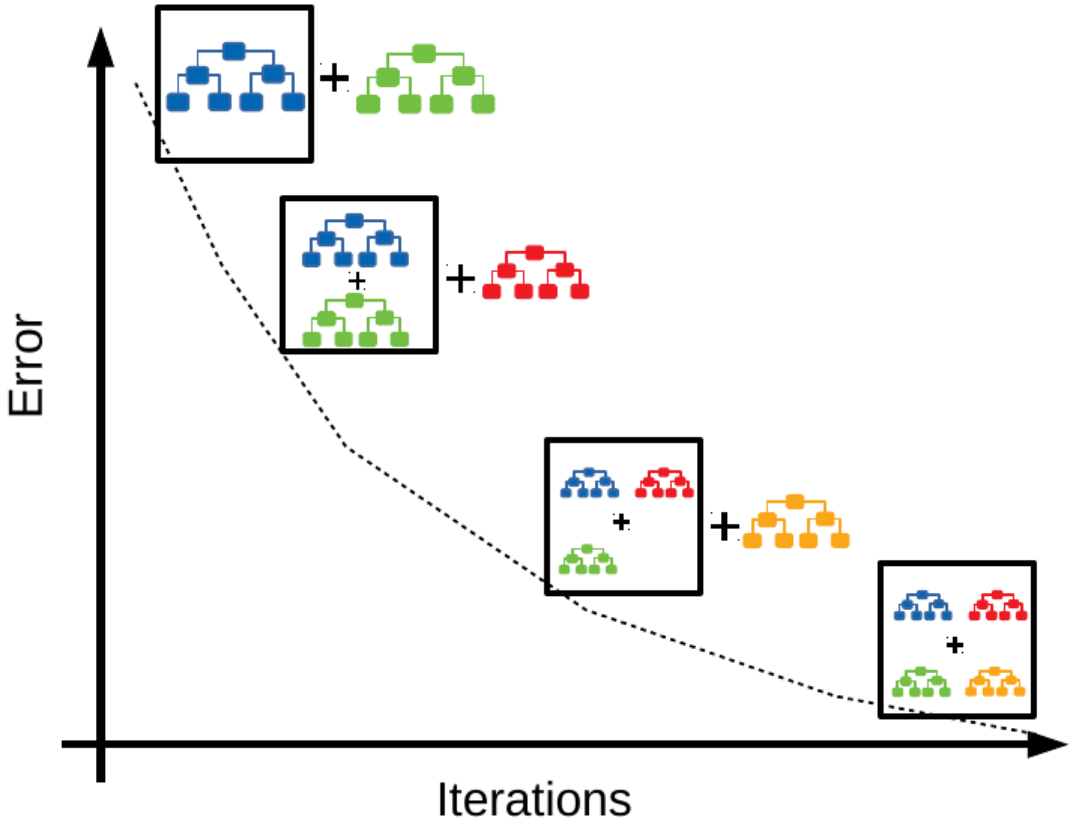
روش Gradient Boosting Machines

- در نهایت خروجی مدل‌های یادگیرنده با هم جمع می‌شوند.
- باید توجه داشت که در GBM بر خلاف RF تعداد زیادی درخت کم عمق استفاده می‌شود.
- انتخاب بهترین تعداد درخت‌ها (انتخاب تعداد مراحل) یکی از هایپرپارامترهای مهم این الگوریتم است.
- روش معروف و قدرتمندXGBoost از خانواده GBM ها است که امروزه در بسیاری از مسابقات با اختلاف بهترین الگوریتم شناخته می‌شود. این روش تعداد زیادی بهینه‌سازی‌های الگوریتمیکی و اصلاح‌شده را به الگوریتم GBM اضافه کرده تا قابلیت توزیع‌شدگی بسیار بیشتری داشته باشد.
- الگوریتم LightGBM یکی دیگر از روش‌های مبتنی‌بر GBM است که از نظر محاسبات، حافظه و زمان اجرا از آن بهینه‌تر است و قادر به انجام یادگیری موازی بهینه است.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

روش Gradient Boosting Machines



GBM Training

Input: A base regression algorithm - I, number of iterations T, the training set, $S = \{(x_i, y_i)\}_{i=1}^{m}$, and a differentiable loss function $L(y, F(x))$

1: Initialize model with a constant value: $F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$

2: $j \leftarrow 1$

3: **repeat**

4: For $i = 1, \ldots, m$, compute pseudo-residuals:
$r_{ij} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{j-1}(x)}$

5: Construct regression model h_j using I using the training set $\{(x_i, r_{ij})\}_{i=1}^{m}$.

6: Find multiplier γ_j by performing line search on the following one-dimensional optimization problem:
$\gamma_j = \arg\min_{\gamma} \sum_{i=1}^{m} L\left(y_i, F_{j-1}(x_i) + \gamma h_j(x_i)\right)$.

7: Update the model: $F_j(x) = F_{j-1}(x) + \gamma_j h_j(x)$.

8: $j \leftarrow j + 1$

9: **until** $j > T$

10: Output $F_M(x)$

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

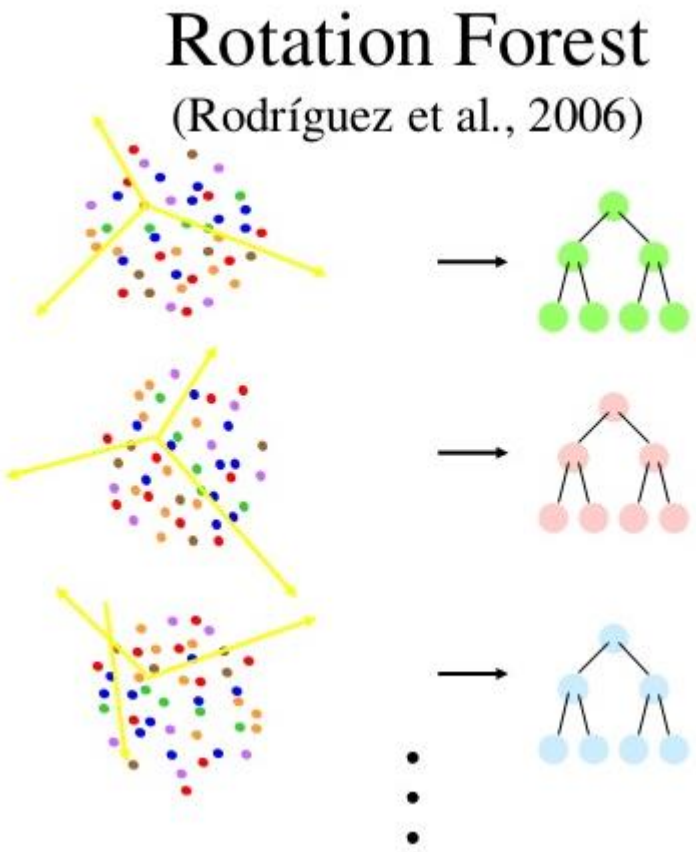
روش Rotation Forest

- این روش تنوع و گوناگونی روی درخت‌های تصمیم را با آموزش هر یادگیرنده روی کل داده‌ها البته روی یک فضای ویژگی چرخش‌یافته ایجاد می‌کند.
- زمانی‌که داده‌های آموزشی برای دسته‌بند پایه ساخته می‌شود، ویژگی‌ها به صورت تصادفی در K زیرمجموعه تقسیم می‌شوند و روی هر زیرمجموعه الگوریتم PCA اعمال می‌گردد.
- این تبدیل تضمین می‌دهد اولین عنصر اصلی دارای بیشترین پراکندگی است.
- اگر چه این روش اغلب نتایجی بهتر از Random Forest دارد، اما دارای دو مشکل اساسی است:
  - محاسبه PCA کاری پرهزینه و زمان‌بر است.
  - تفسیرپزیری هر درخت از بین می‌رود.

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش های یادگیری گروهی

روش Rotation Forest



Rotation Forest

Input: I (a base inducer), S (the original training set), T (number of iterations), and K (number of subsets),

1: **for** $i = 1$ to T **do**
2: Split the feature set into K subsets: $F_{i,j}$ (for j=1..K)
3: **for** $j = 1$ to K **do**
4: Let $S_{i,j}$ be the data set S for the for the features in $F_{i,j}$
5: Eliminate from $S_{i,j}$ a random subset of classes
6: Select a bootstrap sample from $S_{i,j}$ of size 75% of the number of objects in $S_{i,j}$. Denote the new set by $S'_{i,j}$
7: Apply PCA on $S'_{i,j}$ to obtain the coefficients in a matrix $C_{i,j}$
8: **end for**
9: Arrange the $C_{i,j}$, for j = 1 to K in a rotation matrix R_i as in the equation:

$$R_i = \begin{bmatrix} a_{i,1}^{(1)}, a_{i,1}^{(2)}, \ldots, a_{i,1}^{(M_1)} & [0] & \ldots & [0] \\ [0] & a_{i,2}^{(1)}, a_{i,2}^{(2)}, \ldots, a_{i,2}^{(M_2)} & \ldots & [0] \\ \ldots & \ldots & \ldots & \ldots \\ [0] & [0] & \ldots & a_{i,k}^{(1)}, a_{i,k}^{(2)}, \ldots, a_{i,k}^{(Mk)} \end{bmatrix}$$

10: Construct R_i^a by rearranging the columns of R_i so as to match the order of features in F
11: **end for**
12: Build classifier M_i using (SR_i^a, X) as the training set

تولید محتوا: سیدفواد سید ابوترابی
daychegroup
daychegroup
گروه دایکه | dayche.com