یادگیری داده های جریانی
Data Stream Learning

عادل قاضی خانی
هیئت علمی گروه کامپیوتر دانشگاه امام رضا (ع)

تعریف

- داده هایی با طول بی نهایت مربوط به موضوعاتی که همیشگی هستند. مثلاً تراکنشهای کارت اعتباری، آب و هوا و ...
- گونه ای از داده های حجیم

ویژگی های داده های جریانی

- **جریان دائمی**
- **پویایی (تغییر مفهوم)**
- ...

تغییر مفهوم

- در داده های جریانی مدل یادگیری به مرور زمان تغییر می‌کند. یعنی آنچه یادگیری شده بعد از مدتی دیگر قابل استفاده نیست و باید بروزرسانی شود (یادگیری)

یادآوری

- **قانون بیز** که مبنای مدل یادگیری طبقه‌بندها است

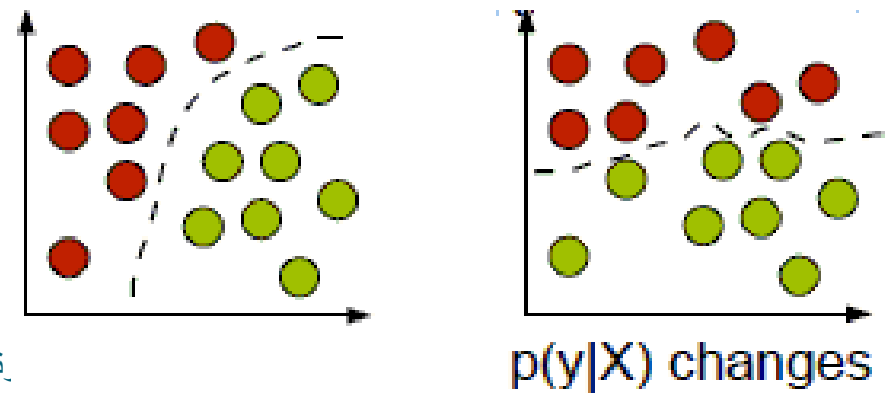
$$P(y/X) = \frac{p(X/y)P(y)}{p(X)}$$

تغییر مفهوم

- **قانون بیز:**

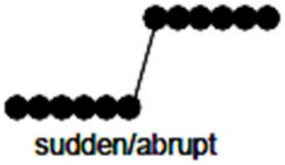
$$P(y/X)=\frac{p(X/y)P(y)}{p(X)}$$

- **تغییر مفهوم:**

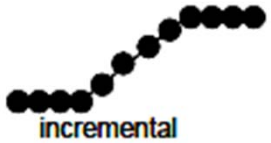


انواع تغییر مفهوم

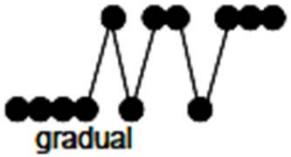
- ناگهانی



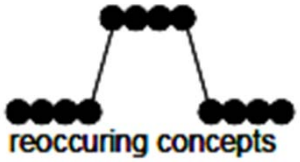
- افزایشی



- تدریجی



- چرخشی



الزامات یادگیری داده های جریانی

- یادگیری بصورت برخط یا افزایشی (تکراری)
- شناسایی و تطبیق با تغییر مفهوم (فعال[1]،غیرفعال[2])
- محدودیت زمانی و حافظه پردازش
- فراموشی[3]

1. Active
2. Passive
3. Forgetting

یادگیری برخط و افزایشی

- در یادگیری برخط[1] در هر لحظه با یک نمونه آموزش مدل انجام می‌گیرد
- در یادگیری افزایشی[2] آموزش مدل با تعدادی نمونه انجام می گیرد

1-online learning
2-incremental learning

یادگیری برخط – روند کلی

- **با دریافت یک نمونه کارهای ذیل انجام می شود:**

- پیش‌بینی
- محاسبه خطا
- بروزرسانی مدل یادگیری

- یک گذر* بر روی داده ها

*** one pass (against multi pass)**

یادگیری برخط – مثال

Algorithm 1. Adaptive λ-Perceptron.

Require: $\eta > 0, \alpha > 0, \lambda(1) \in [0,1), G_\lambda^- < 0, G_\lambda^+ > 0, \lambda^+ \in (\lambda^-, 1)$, $\lambda^- \in [0, \lambda^+)$

Set: $\beta_j(1) = 0, G_{\beta_j}(0) = 0, G_\lambda \beta_j(1) = 0, \forall j = 0, \ldots, d$, $B(0) = 0$, $n(0) = 0$

for $t = 1, 2, \ldots$ **do**

// **Classify**

Predict class label of $x(t)$

Receive true class label $y(t)$

$n(t) = 1 + \lambda(t) n(t-1)$

// **Update** $\beta_j(t)$

for $j = 0, 1, \ldots, d$ **do**

$G_{\beta_j}(t) = \frac{\partial e(\beta(t), t)}{\partial \beta_j(t)} + \lambda G_{\beta_j}(t-1)$

$\beta_j(t+1) = \beta_j(t) - \eta G_{\beta_j}(t)$

end for

// **Update** $\lambda(t)$

$G_\lambda e(t) = \sum_{j=0}^{d} \frac{\partial e(\beta(t), t)}{\partial \beta_j(t)} G_\lambda \beta_j(t)$

if $G_\lambda e(t) > G_\lambda^+$ or $G_\lambda e(t) < G_\lambda^-$ **then**

$\lambda(t+1) = [\lambda(t) - \alpha \operatorname{sign}(G_\lambda e(t))]_{\lambda^-}^{\lambda^+}$

else

$\lambda(t+1) = \lambda(t)$

end if

//**Compute** $G_\lambda \beta_j(t+1)$

for $j = 0, 1, \ldots, d$ **do**

$B_j(t) = \frac{\partial^2 e(\beta(t), t)}{\partial \lambda \partial \beta_j(t)} + G_{\beta_j}(t-1) + \lambda B_j(t-1)$

$G_\lambda \beta_j(t+1) = G_\lambda \beta_j(t) - \frac{\eta}{n(t)} B_j(t)$,

end for

end for

- پیش‌بینی
- محاسبه خطا
- بروزرسانی مدل

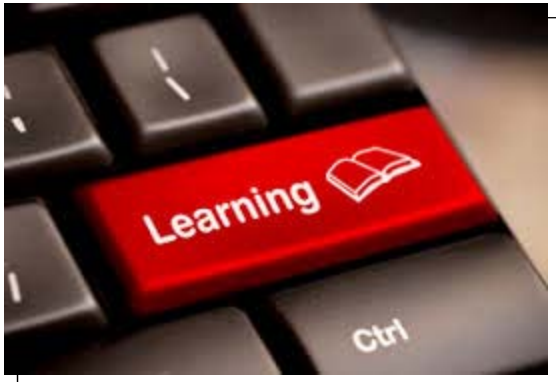
Pavlidis, N. G., Tasoulis, D. K., Adams, N. M., & Hand, D. J. **(2011)**. Landa perceptron: An adaptive classifier for data streams

یادگیری برخط – مثال

- درخت تصمیم *VFDT(Very Fast Decision Tree)*

- در این درخت تصمیم ، اطلاعاتی از داده های ورودی ذخیره می شود سپس با استفاده از آنها معیار انتخاب صفت اندازه گیری شده و بهترین صفات انتخاب می شود

G. Hulten, L. Spencer, P. Domingos, Mining time-changing data streams, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'01), 2001, pp. 97–106.



یادگیری افزایشی – روند کلی

- **با دریافت تعدادی نمونه کارهای ذیل انجام می شود:**

- بروزرسانی مدل یادگیری با بخشی از نمونه ها (آموزشی)
- محاسبه خطای مدل با بخشی از نمونه ها (آزمایشی یا تست)

- چند گذر* بر روی داده ها

یادگیری افزایشی

– مثال استفاده از تعدادی نمونه

- بروزرسانی مدل
- محاسبه خطای مدل

Input: For each dataset $\mathcal{D}^t$ $t = 1,2, \ldots$
Training data $\{x^t(i) \in X; y^t(i) \in Y = \{1, \ldots, c\}\}$, $i = 1, \ldots, m^t$
Supervised learning algorithm **BaseClassifier**
Sigmoid parameters a (slope) and b (infliction point)
Do for $t = 1,2, \ldots$

If $t = 1$, **Initialize** $D^1(i) = w^t(i) = 1/m^1, \forall i,$ (1)
Go to step 3. **Endif**

1. Compute error of the existing ensemble on new data

$$E^t = \Sigma_{i=1}^{mt} 1/m^t . [\![H^{t-1}(x^t(i)) \neq y^t(i)]\!] \quad (2)$$

2. Update and normalize instance weights

$$w_i^t = \frac{1}{m^t} . \begin{cases} E^t, & H^{t-1}(x^t(i)) = y^t(i) \\ 1, & otherwise \end{cases} \quad (3)$$

Set $D^t = w^t / \Sigma_{i=1}^{mt} w^t(i) \Rightarrow D^t$ is a distribution (4)

3. Call **BaseClassifier** with $\mathcal{D}^t$, obtain $h^t: X \rightarrow Y$
4. Evaluate all existing classifiers on new data $\mathcal{D}^t$

$$\varepsilon_k^t = \Sigma_{i=1}^{m^t} D^t(i) [\![h_k(x^t(i)) \neq y^t(i)]\!] \text{ for } k = 1, \ldots, t \quad (5)$$

If $\varepsilon_{k=t}^t > 1/2$, generate a new h_t.
If $\varepsilon_{k<t}^t > 1/2$, set $\varepsilon_k^t = 1/2$,

$$\beta_k^t = \varepsilon_k^t / (1 - \varepsilon_k^t), \text{ for } k = 1, \ldots, t \rightarrow 0 \leq \beta_k^t \leq 1 \quad (6)$$

5. Compute the weighted average of all normalized errors for k^{th} classifier h_k: For $a,b \in \mathbb{R}$

$$\omega_k^t = 1/(1 + e^{-a(t-k-b)}), \omega_k^t = \omega_k^t / \Sigma_{j=0}^{t-k} \omega_k^{t-j} \quad (7)$$

$$\bar{\beta}_k^t = \Sigma_{j=0}^{t-k} \omega_k^{t-j} \beta_k^{t-j}, \text{ for } k = 1, \ldots, t \quad (8)$$

6. Calculate classifier voting weights

$$W_k^t = \log(1/\bar{\beta}_k^t), \quad for\ k = 1, \ldots, t \quad (9)$$

7. Obtain the final hypothesis

$$H^t(x^t(i)) = \arg\max_c \Sigma_k W_k^t . [\![h_k(x^t(i)) = c]\!] \quad (10)$$

Elwell, R., & Polikar, R .**(2011)** . Incremental Learning of Concept Drift in Nonstationary Environments. Neural Networks, IEEE Transactions on, 22(10),

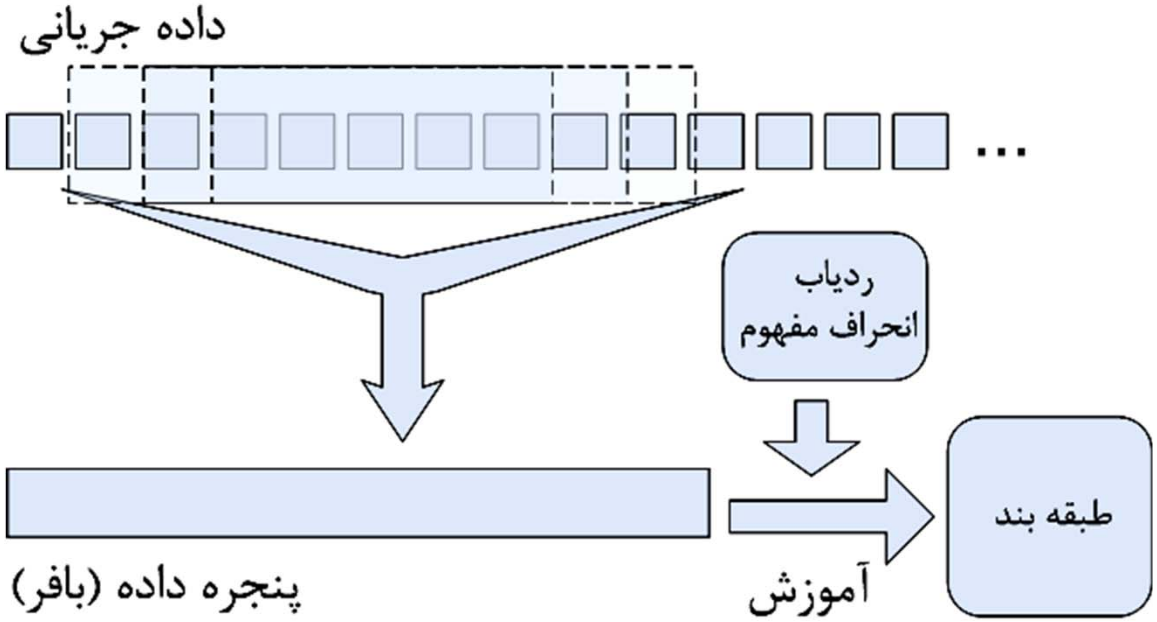
شناسایی و تطبیق با تغییر مفهوم

- **فعال:** شناسایی تغییر مفهوم و سپس بروزرسانی مدل یادگیری با چند نمونه داده

- **غیرفعال:** عدم شناسایی تغیر مفهوم، برخورد دائمی با تغییر مفهوم، رویکرد اصلی وزن‌دهی. داده های جدیدتر وزن بیشتری دارند

شناسایی و تطبیق با تغییر مفهوم

- **فعال:** شناسایی تغییر مفهوم و سپس بروزرسانی مدل یادگیری با چند نمونه داده



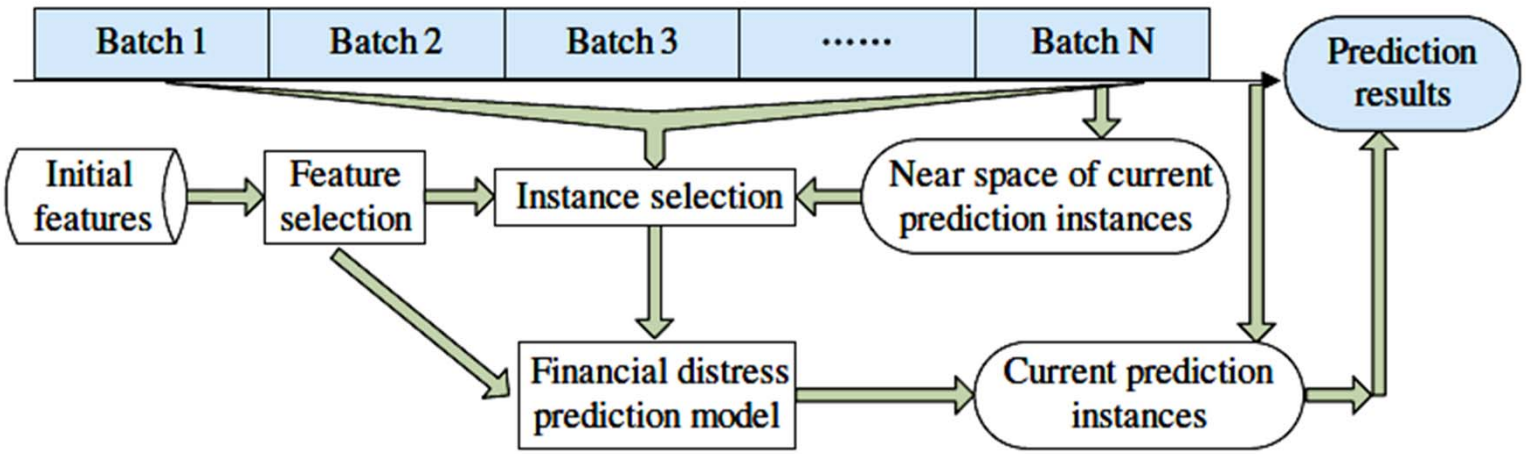
شناسایی و تطبیق با تغییر مفهوم

- **روش فعال**
- **مثال الگوریتم ADWIN**
- در این روش با توجه به اختلاف بین میانگین داده ها در دو پنجره اخیر از داده ها تغییر مفهوم شناسایی می شود
- اندازه پنجره می تواند ثابت یا متغیر در نظر گرفته شود

Bifet, Albert, and Ricard Gavalda. "Learning from time-changing data with adaptive windowing." In Proceedings of the 2007 SIAM international conference on data mining, pp. 443-448. Society for Industrial and Applied Mathematics, 2007.

شناسایی و تطبیق با تغییر مفهوم

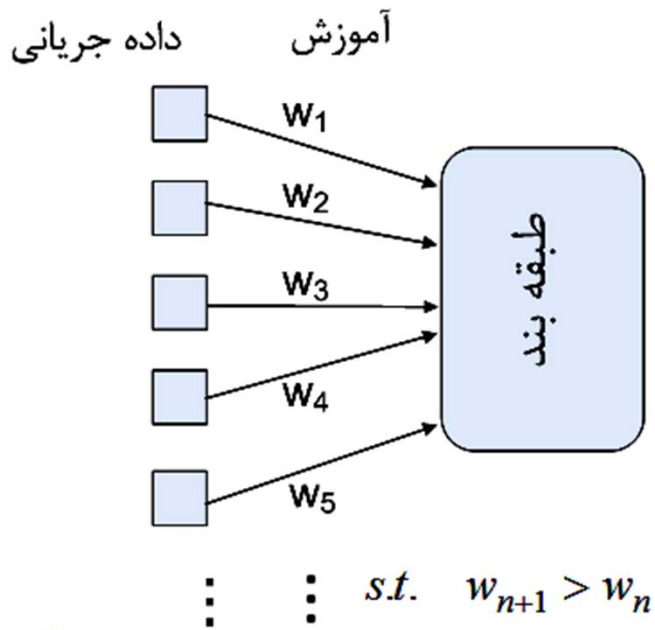
- فعال – مثال



Sun, J., & Li, H. **(2011)**. Dynamic financial distress prediction using instance selection for the disposal of concept drift. Expert Systems with Applications, 38(3), 2566-2576.

شناسایی و تطبیق با تغییر مفهوم

- **غیرفعال:** عدم شناسایی تغیر مفهوم، برخورد دائمی با تغییر مفهوم، رویکرد اصلی وزن‌دهی. داده های جدیدتر وزن بیشتری دارند



$$\vdots \quad \vdots \quad s.t. \quad w_{n+1} > w_n$$

شناسایی و تطبیق با تغییر مفهوم

- **غیرفعال – مثال**

- **رابطه خطای پرسپترون بدون وزن دهی**

$$\mathcal{E}(\beta,t)=\sum_{i=1}^{t} e(\beta,i)=\frac{1}{2}\sum_{i=1}^{t}(y(i)-\varphi(\beta^{\mathsf{T}}z(i)))^2$$

- **رابطه خطای پرسپترون با وزن دهی**

$$\mathcal{E}(\beta,t)=\sum_{i=1}^{t}\lambda^{t-i}e(\beta,i)=e(\beta,t)+\lambda\mathcal{E}(\beta,t-1).$$

Pavlidis, N. G., Tasoulis, D. K., Adams, N. M., & Hand, D. J. **(2011)**. Landa perceptron: An adaptive classifier for data streams, Pattern Recognition, 44(1), 78-96 .

محدودیت زمانی و حافظه پردازش

- در یادگیری داده های جریانی در حالت **برخط** و **افزایشی** محدودیت حافظه و زمان مورد توجه است

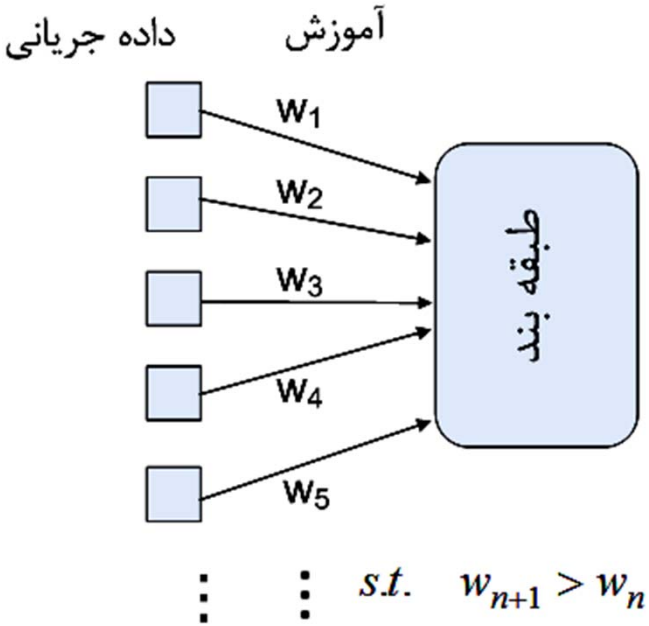
فراموشی

- بدلیل پویایی داده های جریانی فراموشی یک ویژگی الگوریتمهای یادگیری است

- **تدریجی**: عمدتاً از طریق وزن‌دهی به داده‌ها
- **ناگهانی**: با تعریف یک بافر برای ذخیره داده‌های اخیر . در زمان تغییر مفهوم فراموشی ناگهانی داده های قبلی اتفاق می افتد

فراموشی

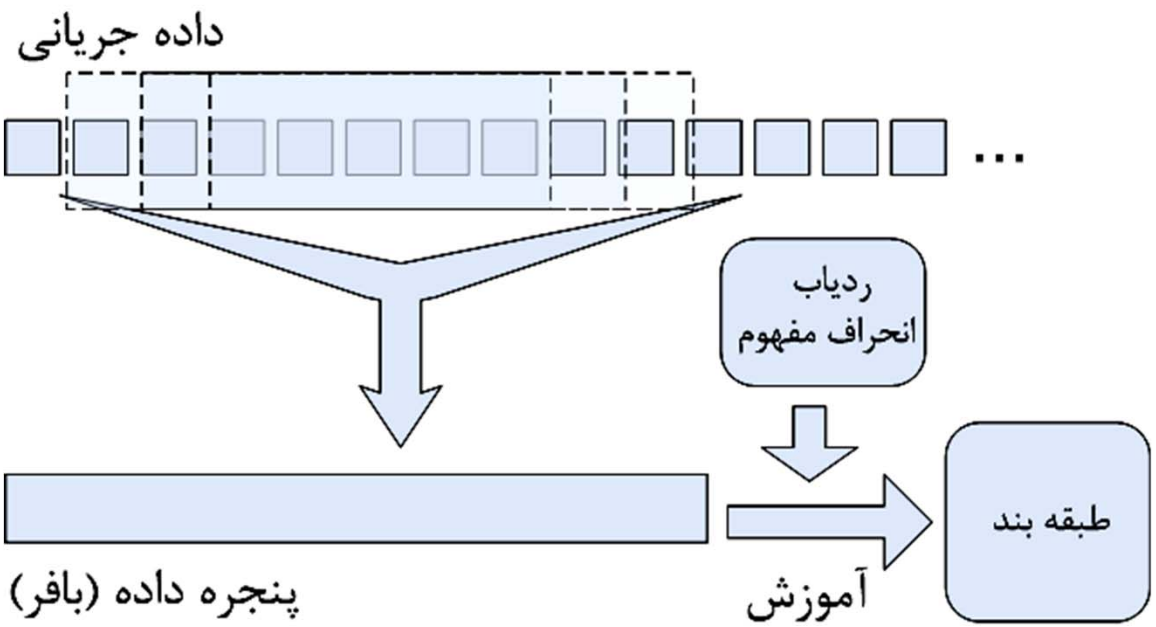
- **تدریجی**: عمدتاً از طریق وزن‌دهی به داده‌ها



$s.t. \quad w_{n+1} > w_n$

فراموشی

- **ناگهانی**: با تعریف یک بافر برای ذخیره داده‌های اخیر . در زمان تغییر مفهوم فراموشی ناگهانی داده های قبلی اتفاق می افتد

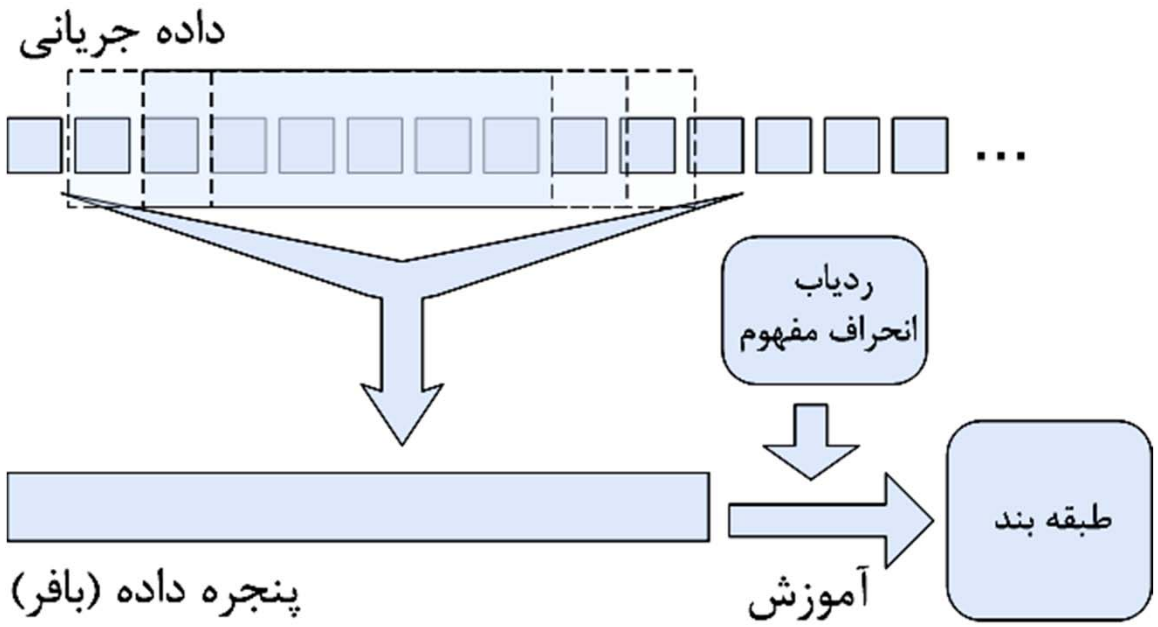


نوع یادگیری

- طبقه‌بندی و خوشه‌بندی
- یادگیری تک و ترکیبی

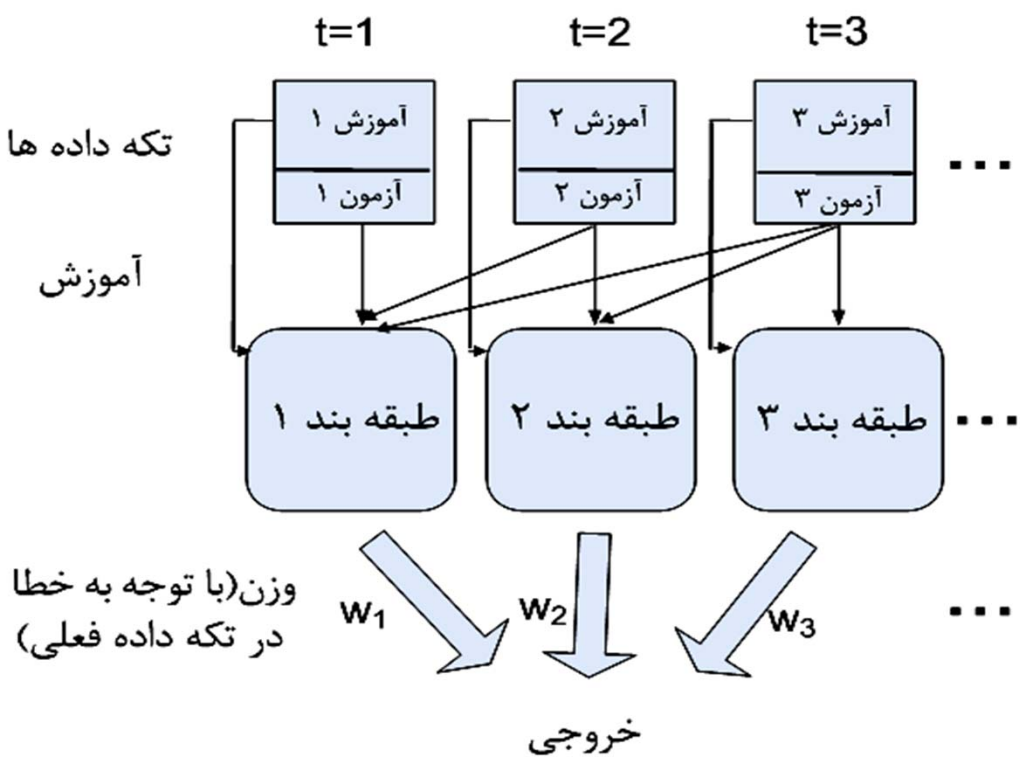
نوع یادگیری

- **طبقه‌بندی یادگیری تک**



نوع یادگیری - مثال

- **طبقه‌بندی یادگیری ترکیبی**



جمع‌بندی

- داده های جریانی گونه ای از داده های حجیم هستند
- ویژگی اساسی آنها تغییر مفهوم
- یادگیری با محدودیتهای خاص