

# ترجمه کامل مقاله:

## An Overview of the Supervised Machine Learning Methods

### (Abstract) چکیده

در دهه گذشته تعداد زیادی از روش‌های یادگیری نظارت شده در حوزه یادگیری ماشین معرفی شده‌اند. یادگیری نظارت شده به یکی از موضوعات اصلی تحقیقات در یادگیری ماشین تبدیل شده است و بسیاری از تکنیک‌های آن برای پردازش و تحلیل انواع داده‌ها به کار گرفته می‌شوند.

یکی از ویژگی‌های مهم یادگیری نظارت شده، استفاده از داده‌های برچسب‌خورده است که شامل برچسب‌هایی برای فرایند طبقه‌بندی هستند.

این مقاله جنبه‌های بنیادی چند روش نظارت شده را خلاصه می‌کند و هدف اصلی آن، ارائه یک مرور کلی از یادگیری ماشین و معرفی تکنیک‌های آن است.

### (Introduction) مقدمه

یادگیری ماشین حوزه‌ای گسترده است که در فناوری اطلاعات، آمار، احتمال، هوش مصنوعی، روان‌شناسی، نوروپیولوژی و بسیاری از علوم دیگر استفاده می‌شود.

در یادگیری ماشین، حل مسائل با ساختن مدلی انجام می‌شود که نماینده مناسبی از مجموعه داده انتخابی باشد.

با پیشرفت فناوری‌های پردازشی و داده‌های عظیم (Big Data)، یادگیری ماشین امروزه بسیار متفاوت از گذشته است. امروزه الگوریتم‌های متعددی توسعه یافته، بهبود پیدا کرده و قادر به انجام محاسبات پیچیده روی داده‌های بزرگ هستند.

برنامه‌نویسی تطبیقی نیز بسیار محبوب شده است؛ جایی که برنامه‌ها قادر به تشخیص الگوهای یادگیری از تجربه، استخراج اطلاعات جدید از داده‌ها و بهینه‌سازی دقت و کارایی خروجی خود هستند.

### انواع یادگیری ماشین

بر اساس نوع خروجی مورد انتظار، الگوریتم‌های یادگیری ماشین به دسته‌های زیر تقسیم می‌شوند:

#### 1. یادگیری نظارت شده:

ایجاد تابعی که ورودی‌ها را به خروجی‌های مطلوب نگاشت کند. غالباً در مسائل طبقه‌بندی استفاده می‌شود.

#### 2. یادگیری بدون نظارت:

داده‌ها بدون برچسب هستند و هدف پیدا کردن ساختار پنهان میان داده‌هاست.

3. یادگیری نیمه نظارت شده: ترکیبی از داده های برچسب خورده و بدون برچسب.
4. یادگیری تقویتی: الگوریتم با تعامل با محیط و دریافت پاداش یاد می گیرد.
5. یادگیری انتقالی / استنتاجی (Transduction): بدون ساختتابع صریح، خروجی های جدید را از ورودی های جدید و قدیمی حدس می زند.
6. یادگیری برای یادگیری (Learning to Learn): الگوریتم از تجربیات قبلی، تمایلات استنتاجی خود را یاد می گیرد.

## کارهای مرتبط (Related Work)

مقالات متعددی تکنیک های یادگیری ماشین و طبقه بندی را بررسی کرده اند.  
از جمله:

- مقاله Caruana & Mizil که ۱۰ روش یادگیری نظارت شده (از جمله SVM، شبکه عصبی، لجستیک، بیز ساده و ... ) را مقایسه می کند.
- پژوهش هایی درباره طبقه بندی اسناد متنی، تحلیل داده های پژوهشی، و مقایسه دقت، سرعت و پیچیدگی الگوریتم ها.
- کتاب ها و فصل هایی که درباره KNN، درخت تصمیم و سایر ابزارهای یادگیری نظارت شده توضیح می دهند.

## پس زمینه: یادگیری نظارت شده (Background: Supervised Learning)

فرایند یادگیری نظارت شده شامل دو مرحله است:

1. آموزش (Training): ویژگی ها توسط الگوریتم آموخته شده و مدل ساخته می شود.
2. آزمون (Testing): مدل روی داده های جدید اعمال شده و خروجی پیش بینی می شود.

الگوریتم، داده های ورودی را همراه با برچسب های صحیح دریافت می کند و با مقایسه خروجی و مقدار واقعی، خطای کاهش می دهد.

### دو نوع اصلی وظایف نظارت شده

- طبقه بندی (Classification): خروجی گسسته است.
- رگرسیون (Regression): خروجی عددی و پیوسته است.

یادگیری نظارت شده بیشترین کاربرد را در شبکه های عصبی، درخت تصمیم و کاربردهای پیش بینی داده های تاریخی دارد (مثل پیش بینی نوع گل زنبق).

## الگوریتم های یادگیری نظارت شده (Supervised Learning Algorithms)

### (1) درخت تصمیم (Decision Tree)

- درخت شامل گره ریشه، گره های داخلی و برگ هاست.
- هر گره یک آزمون روی ویژگی ها انجام می دهد.
- نمونه با عبور از گره ها به برگ مناسب می رسد.
- می تواند ویژگی های عددی و رده ای را پردازش کند.
- ساده، قابل فهم، اما مستعد بیش برازش.

### (2) رگرسیون خطی (Linear Regression)

- هدف: یافتن رابطه خطی بین متغیرهای ورودی و خروجی.
- خروجی پیوسته است.
- مدل با کمینه سازیتابع هزینه، خط مناسب را پیدا می کند.

### (3) بیز ساده (Naïve Bayes)

- مبتنی بر قضیه بیز.
- فرض استقلال ویژگی ها دارد.
- سریع، مقاوم در برابر نویز، و مناسب طبقه بندی متنی.

### (4) رگرسیون لجستیک (Logistic Regression)

- یک طبقه بند تمایزی (Discriminative) است.
- از تابع سیگموید استفاده می کند:

$$\frac{1}{z - e + 1} = g(z)$$

- خروجی بین ۰ و ۱ است و احتمال رخداد یک کلاس را نشان می دهد.

#### (۵) نزدیک‌ترین همسایه‌ها(KNN)

- نمونه جدید با توجه به  $K$  همسایه نزدیک خود طبقه‌بندی می‌شود.
- محاسبات زیاد دارد اما ساده و دقیق است.

#### (۶) ماشین بردار پشتیبان(SVM)

- به دنبال یافتن بهترین ابرصفحه است که دو کلاس را از هم جدا کند.
- برای داده‌های با ابعاد بالا بسیار مناسب است.
- با کرنل‌ها قابلیت تفکیک داده‌های غیرخطی را دارد.

#### (۷) جنگل تصادفی(Random Forest)

- مجموعه‌ای از درخت‌های تصمیم است که با داده‌ها و ویژگی‌های تصادفی ساخته می‌شوند.
- دقت بالا و مقاومت خوب در برابر بیش‌بازش دارد.

#### AdaBoost(۸)

- چندین مدل ضعیف را با وزن‌دهی متفاوت ترکیب می‌کند.
- نمونه‌های اشتباه وزن بیشتری می‌گیرند تا مدل‌های بعدی روی آن‌ها تمرکز کنند.

#### Gradient Boosting(۹)

- مدل‌ها به صورت مرحله‌ای ساخته می‌شوند و هر مدل خطای مدل قبلی را اصلاح می‌کند.
- دقت بسیار بالا، ولی حساس به بیش‌بازش.

---

#### XGBoost(۱۰)

- نسخه سریع‌تر و بهینه‌تر Gradient Boosting.
- دارای Regularization (جلوگیری از بیش‌بازش).
- از بهترین الگوریتم‌ها در رقابت‌های Kaggle.

## ۱۱ (یادگیری عمیق) (Deep Learning)

- شامل شبکه‌های عصبی چندلایه (CNN، RNN و ...).
- قدرت بسیار بالا در تحلیل تصویر، صوت و متن.
- نیازمند داده و توان پردازشی زیاد.

## (Conclusion) (نتیجه‌گیری)

می‌توان نتیجه گرفت که یادگیری نظارت شده یکی از مهم‌ترین روش‌ها در یادگیری ماشین است. روش‌های نظارت شده معمولاً موفق‌تر از روش‌های بدون ناظر هستند زیرا داده‌های برچسب‌خورده معیارهای واضحی برای آموزش مدل فراهم می‌کنند.  
این مقاله مرور کوتاهی از فرایند یادگیری، ساختار الگوریتم‌ها و برخی از روش‌های رایج یادگیری نظارت شده ارائه داد.

با پیشرفت سریع این حوزه، فضای زیادی برای توسعه و بهبود الگوریتم‌ها وجود دارد و انتظار می‌رود عملکرد آن‌ها در آینده حتی بهتر شود.