



COHORT STUDIES & BIOBANKS

MEDICAL DATA SCIENCE

USERS: LAST 7 DAYS USING MEDIAN

یادگیری ماشین

در تحلیل داده‌های کوهورت

از داده‌های چندلایه و پیچیده بیوبانک‌ها تا کشف الگوهای پنهان و پیش‌بینی دقیق پیامدهای سلامت آینده

ارائه‌دهنده
دکتر یاسین فدائی



مطالعات کوهورت و بیوبانکها

سیستمهای چندلایه: فراتر از دادههای ساده



چالشهای تحلیل داده



- ناهمگنی و ابعاد بالا (High-dimensional): تعداد متغیرها بسیار بیشتر از تعداد نمونههاست.
- دادههای گمشده (Missing Data): عدم حضور بیماران در تمام مراحل پیگیری.
- الگوهای غیرخطی: روابط پیچیده بین ژن و محیط که آمار کلاسیک نمی بیند.

فرصت طلایی



استفاده از یادگیری ماشین برای ادغام چندوجهی (Multi-modal Integration) این لایهها و کشف الگوهای پنهانی که در تحلیلهای تکمتغیره قابل مشاهده نیستند.

→ حرکت به سمت پزشکی شخصی سازی شده

لایههای داده در بیوبانک



در بیوبانکها فقط با یک نوع داده روبرو نیستیم، بلکه با یک اکوسیستم پیچیده از اطلاعات مواجهیم که در طول سالها جمعآوری شدهاند.

دادههای بالینی و دموگرافیک

سن، جنسیت، سوابق بیماری، داروها



دادههای امیکس (ژنومی)

توالی یابی DNA، بیان ژن، پروتئومیکس



سبک زندگی و محیط

تغذیه، فعالیت فیزیکی، آلایندهها، وضعیت اجتماعی



تصویربرداری پزشکی

MRI, CT Scan, سونوگرافی



پیامدهای طولی (Follow-up)

رخدادهای مرگومیر، بستری، عود بیماری طی زمان



چه سؤالاتی را می‌توان پاسخ داد؟

یادگیری ماشین فراتر از توصیف داده‌ها عمل می‌کند:

حوزه‌های کاربردی

این پرسش‌ها در حوزه‌های زیر بیشترین کاربرد را دارند:

قلب و عروق

Cardiology



سرطان (Oncology)

تشخیص زودرس و بقا



دیابت و متابولیسم

مدیریت عوارض مزمن



سلامت جمعیت

Population Health



نکته کلیدی:

قدرت اصلی ML در ترکیب داده‌های ناهمگن (Heterogeneous) از این حوزه‌ها نهفته است.

چه کسی در آینده بیمار می‌شود؟

پیش‌بینی رخداد بیماری (Incidence) قبل از بروز علائم بالینی



کدام بیماران ریسک بالاتری دارند؟

طبقه‌بندی خطر (Risk Stratification) برای اولویت‌بندی مراقبت‌ها



کدام الگوها با پیامد مرتبط‌اند؟

کشف تعاملات پیچیده و غیرخطی بین متغیرهای دموگرافیک و ژنتیک



کدام Biomarkerها پیش‌بینی‌کننده‌اند؟

انتخاب ویژگی (Feature Selection) برای یافتن نشانگرهای حیاتی



زیرگروه‌های پنهان جمعیت کدام‌اند؟

خوشه‌بندی بدون نظارت (Clustering) برای یافتن فنوتیپ‌های جدید



یادگیری ماشین: تعریف و انواع مدل‌ها

ابزاری برای تبدیل داده‌های خام به بینش بالینی



دسته‌بندی اصلی مدل‌ها

پرکاربردترین

یادگیری با نظارت (Supervised)

هدف: پیش‌بینی یک خروجی مشخص (Label) بر اساس داده‌های ورودی.

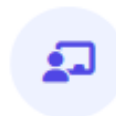
Neural Networks

XGBoost

SVM

Random Forest

Regression



یادگیری بدون نظارت (Unsupervised)

هدف: کشف الگوهای پنهان و گروه‌بندی داده‌ها بدون داشتن برچسب قبلی.

Hierarchical Clustering

PCA (کاهش ابعاد)

K-means Clustering



حیاتی در پزشکی

تحلیل بقا و زمان تا رخداد (Survival)

هدف: پیش‌بینی زمان احتمالی وقوع یک رخداد (مثل مرگ یا عود بیماری).

DeepSurv

Random Survival Forest

Cox PH



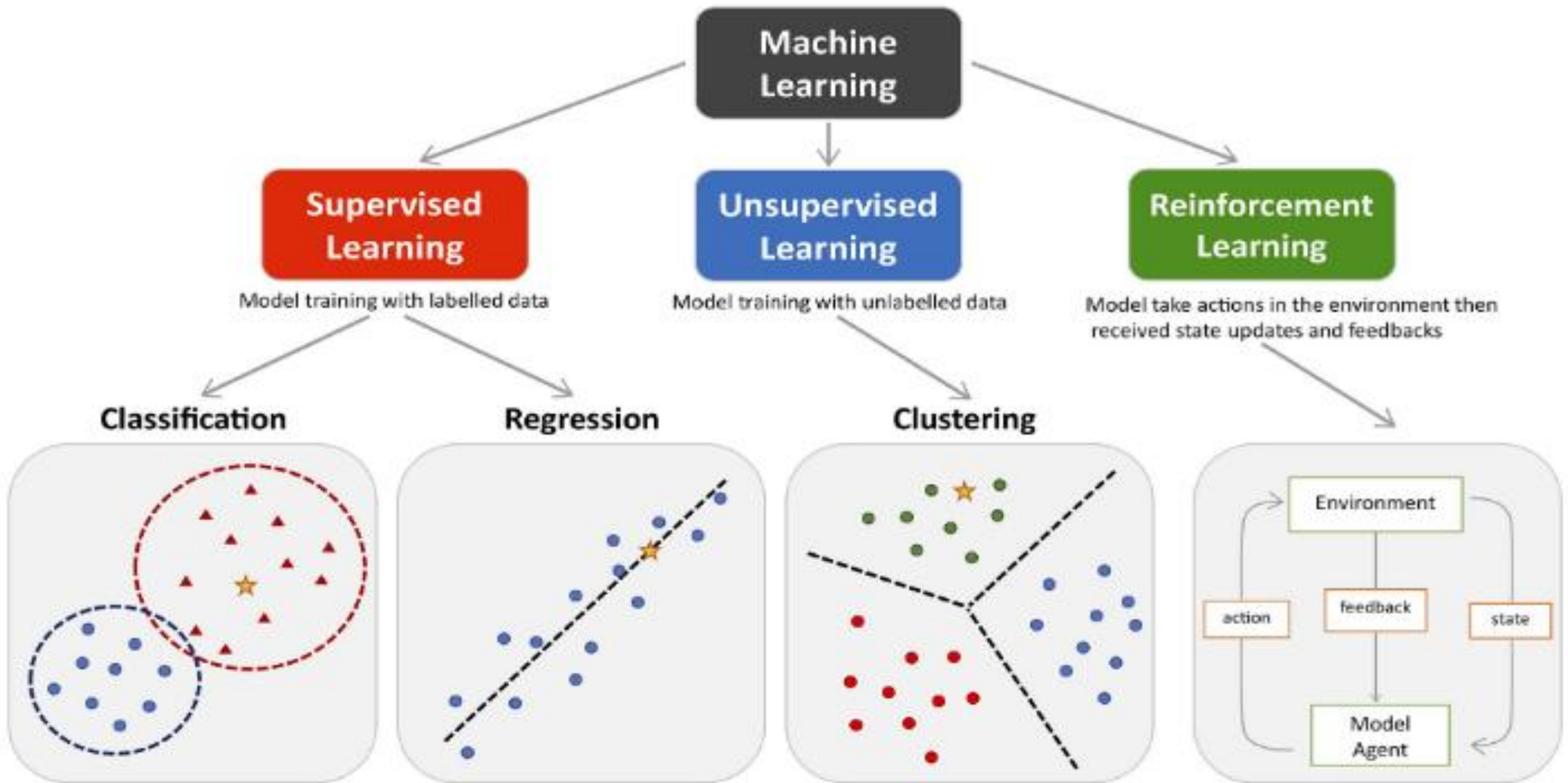
یادگیری ماشین چیست؟

یادگیری ماشین، شاخه‌ای از هوش مصنوعی است که به کامپیوترها امکان می‌دهد بدون برنامه‌نویسی صریح، از **داده‌های گذشته** بیاموزند و الگوها را کشف کنند.

تمثیل بالینی: درست مثل یک پزشک با تجربه که با دیدن هزاران بیمار، علائم ظریف را تشخیص می‌دهد، مدل ML با دیدن هزاران رکورد داده، الگوهای پیچیده بیماری را یاد می‌گیرد.



انواع مدل های یادگیری ماشین



یادگیری با نظارت

Features ← → **Label**

	x1	x2	...	x_n	y
S_1					
S_2					
...					
S_n					

$$X \xrightarrow[\text{قانون}]{R} Y$$



Pipeline استاندارد یادگیری ماشین

مسیر علمی تبدیل داده خام به دانش معتبر





جمع‌بندی و کاربردها

از مدل‌های ریاضی تا بهبود سلامت جامعه

کاربردهای عملی در سلامت



پزشکی شخصی‌سازی شده

انتخاب بهترین درمان و دوز دارویی برای هر فرد بر اساس پروفایل ژنتیکی و سبک زندگی منحصر به فرد او.



غربالگری هوشمند ریسک

شناسایی افراد پرخطر (High-risk) برای بیماری‌های قلبی یا سرطان، سال‌ها قبل از بروز علائم بالینی.



مدیریت سلامت جمعیت

پیش‌بینی بار بیماری‌ها در آینده و تخصیص بهینه منابع بهداشتی در سطح کلان کشور.



طراحی کارآزمایی‌های بالینی

انتخاب هوشمند بیماران برای شرکت در تحقیقات دارویی جهت افزایش شانس موفقیت و کاهش هزینه‌ها.

پیام اصلی

یادگیری ماشین جایگزین آمار کلاسیک نیست، بلکه **مکمل قدرتمندی** برای داده‌های پیچیده و چندلایه است. هدف نهایی، تبدیل داده‌های ذخیره شده در بیوبانک‌ها به ابزارهای **پیش‌بینی بالینی** برای نجات جان انسان‌هاست.

گام‌های بعدی (Next Steps)

بهبود کیفیت داده
استانداردسازی جمع‌آوری داده در بیوبانک‌ها

حاکمیت و اخلاق
حفظ حریم خصوصی و شفافیت مدل‌ها (XAI)

اعتبارسنجی خارجی
تست مدل‌ها روی جمعیت‌های متفاوت

استقرار در کلینیک