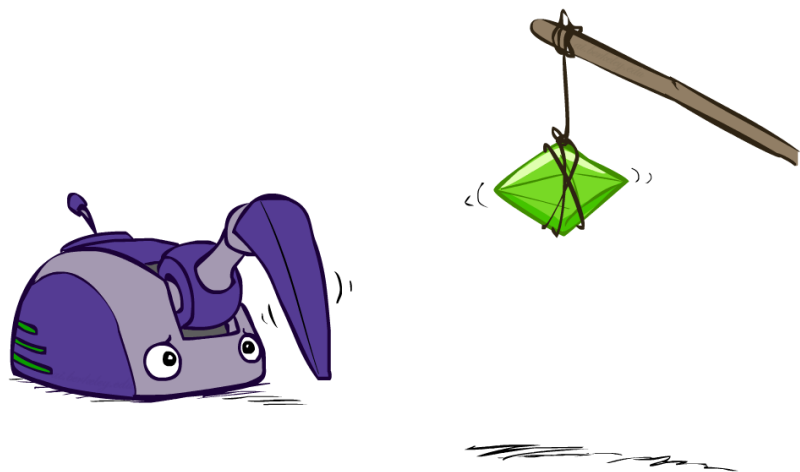


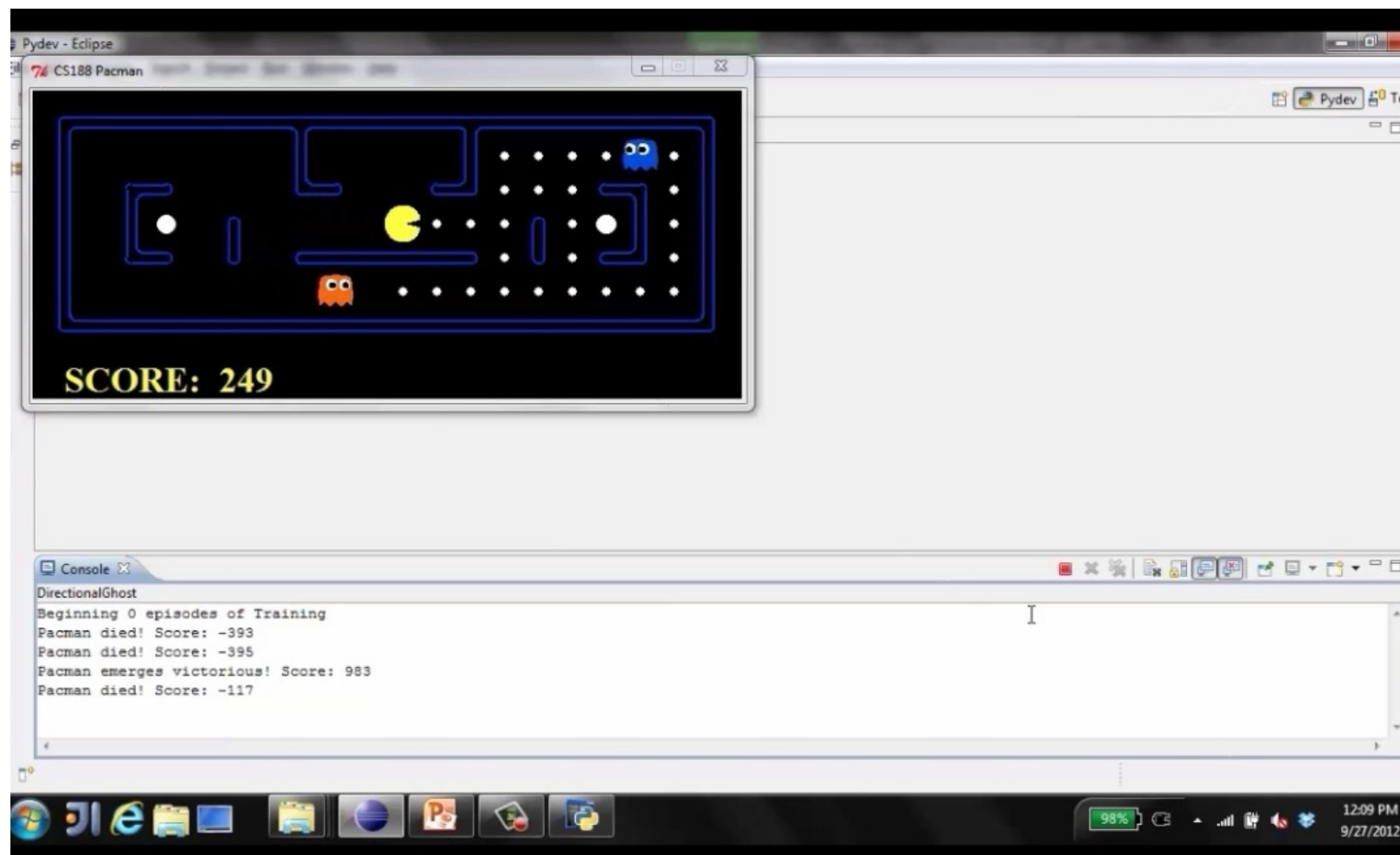
# یادگیری تقویتی (Reinforcement Learning)

1. فرایند تصمیم مارکوف (Markov Decision Processes)
2. الگوریتم تکرار مقدار (Value Iteration)
3. الگوریتم تکرار سیاست (Policy Iteration)
4. یادگیری تقویتی (Reinforcement Learning)

# یادگیری تقویتی (Reinforcement Learning)



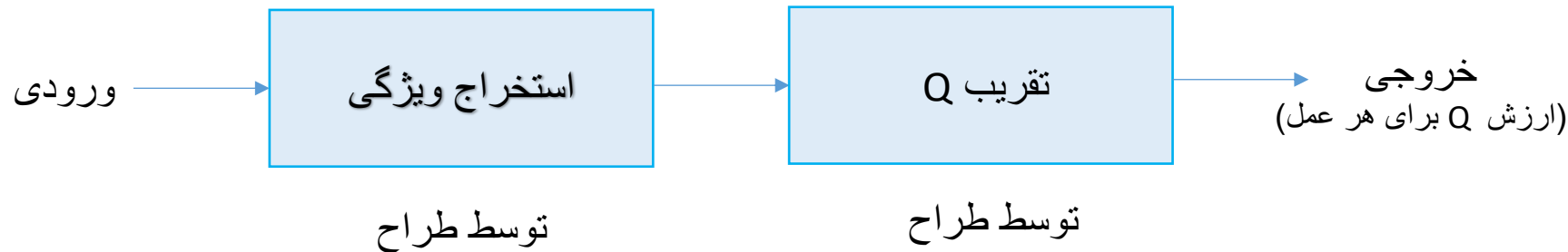
# ویدئوی نمایشی از یادگیری تقریبی Q در بازی پکمن



# روش سنتی (Approximate Q-Learning)

✓ طراح تعدادی ویژگی برای مساله طراحی می کند. ( نیاز به دانش طراح)

✓ طراح، قوانینی را برای استفاده از این ویژگی ها مشخص می کند تا تقریب تابع  $Q$  و انتخاب عمل انجام شود (نیاز به دانش طراح)

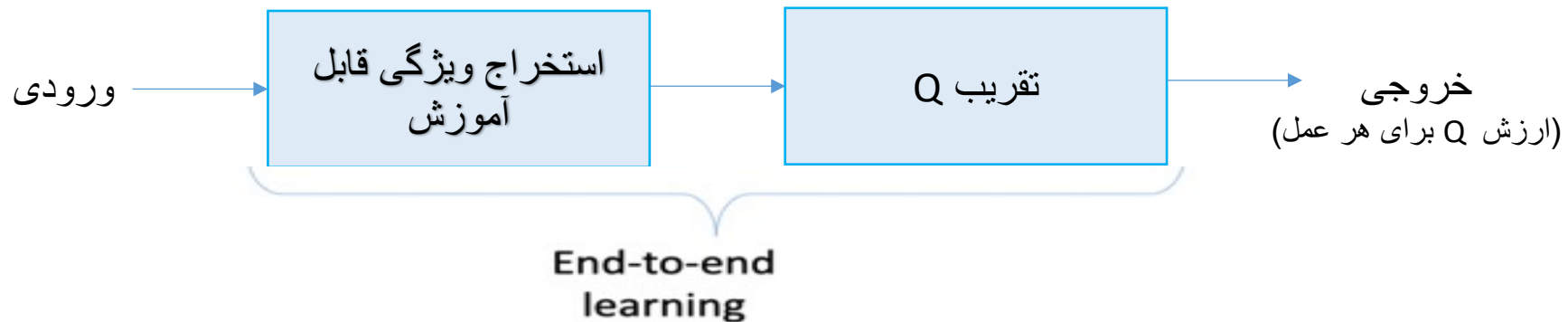


✓ در این روش، تمام تصمیم گیری ها بر پایه ی ویژگی ها و قوانین تعریف شده توسط طراح انجام می شود.

# تقریب Q با یادگیری ماشین

✓ استخراج ویژگی توسط ماشین در یک مرحله یاد گرفته می شود.

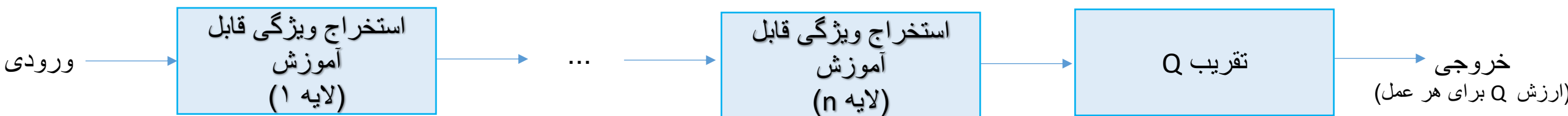
✓ قوانین برای ویژگی ها، توسط ماشین یادگیری و تعیین می شود.



# تقریب Q با یادگیری عمیق

✓ استخراج ویژگی به صورت سلسله مراتبی توسط ماشین یاد گرفته می شود. (هر لایه از روی لایه قبلی ویژگی ها را محاسبه می کند)

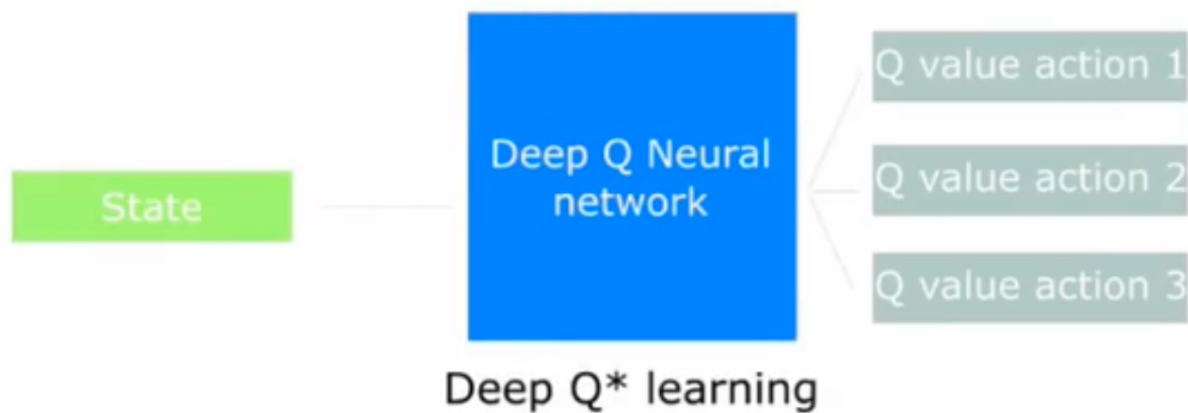
✓ یافتن قوانین برای ویژگی ها، توسط ماشین یادگیری شود.



# Deep Q-Learning

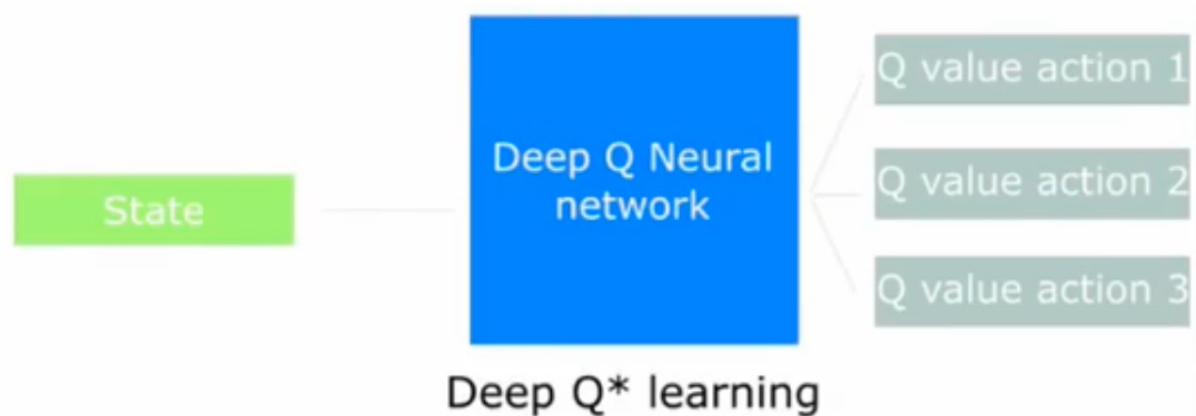


❑ تولید و به روز رسانی یک جدول Q ممکن است در محیط‌هایی با فضای حالت بزرگ ناکارآمد شود.



❑ به جای استفاده از جدول Q، می‌توان یک شبکه‌ی عصبی پیاده‌سازی کرد که یک حالت را دریافت کرده و مقادیر Q را برای هر عمل، بر اساس آن حالت، تقریب بزند.

# Deep Q-Learning



□ به جای استفاده از جدول Q، می توان یک شبکه ی عصبی پیاده سازی کرد که یک حالت را دریافت کرده و مقادیر Q را برای هر عمل، بر اساس آن حالت، تقریب بزند.



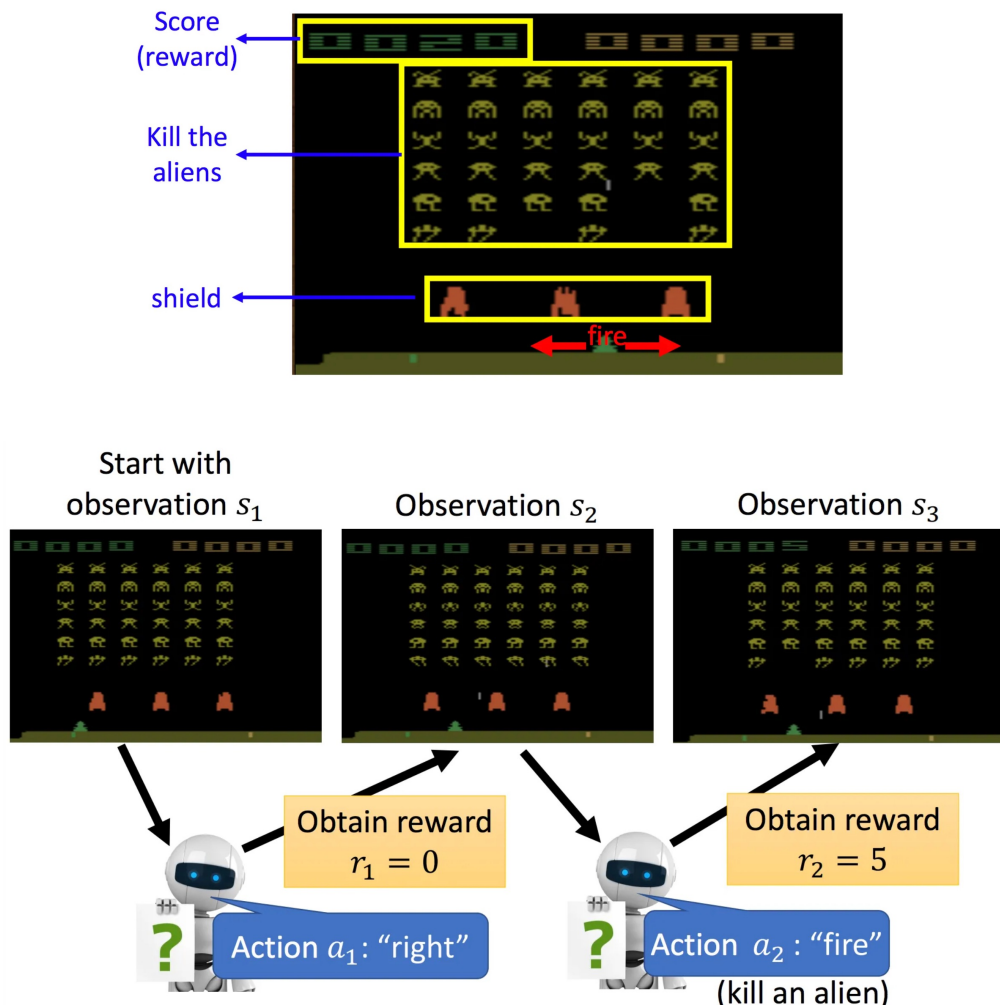


# Deep Q-Learning

## بازی Space Invaders □

▪ محیط دارای ویژگی‌های مانند دشمن، سپر محافظ و امتیاز است که باید از طریق مشاهده و یادگیری بهینه شوند.

- در هر گام، عامل تصویری از محیط را به عنوان **مشاهده** دریافت می‌کند.
- بر اساس این مشاهده، عامل یک **عمل** انتخاب می‌کند (مثلاً حرکت به راست یا شلیک کردن).
- پس از انجام عمل، عامل یک **پاداش** دریافت می‌کند.
- هدف عامل، یادگیری سیاستی است که بتواند بیشترین پاداش ممکن را در طول زمان به دست آورد.



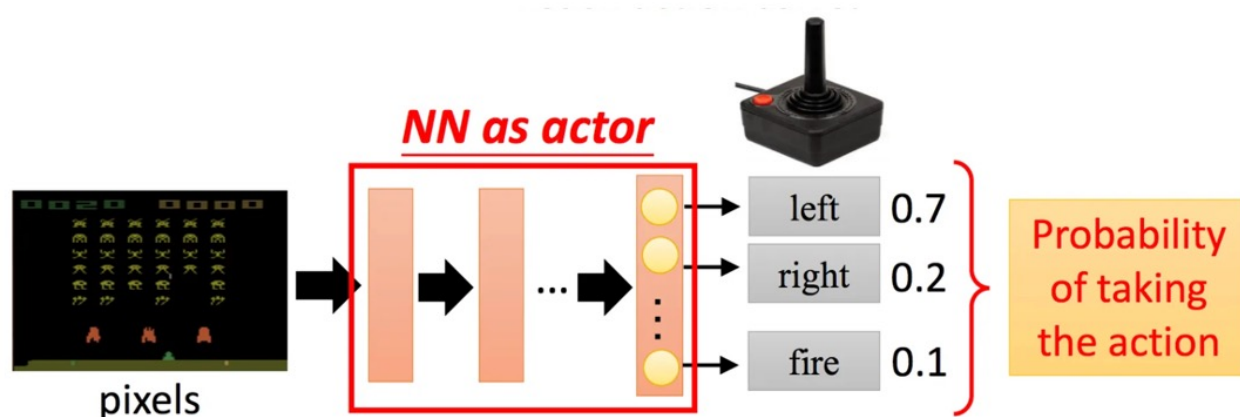
# Deep Q-Learning

## □ بازی Space Invaders:

- ورودی شبکه‌ی عصبی: مشاهده‌ی ماشین که به صورت یک بردار یا ماتریس نمایش داده می‌شود.
- خروجی شبکه‌ی عصبی: یک بردار از عمل‌ها (هر بردار متناظر با یک نورون در لایه‌ی خروجی) است. باید بیشترین مقدار Q از این بردار انتخاب کنیم.

## □ مزیت استفاده از شبکه (عصبی) به جای جدول Q چیست؟

✓ تعمیم‌دهی (Generalization)



## یادگیری ماشین

یادگیری تقویتی

یادگیری با نظارت

یادگیری بدون نظارت

یادگیری نیمه نظارتی

یادگیری خودنظارتی

# یادگیری ماشین (Machine Learning)

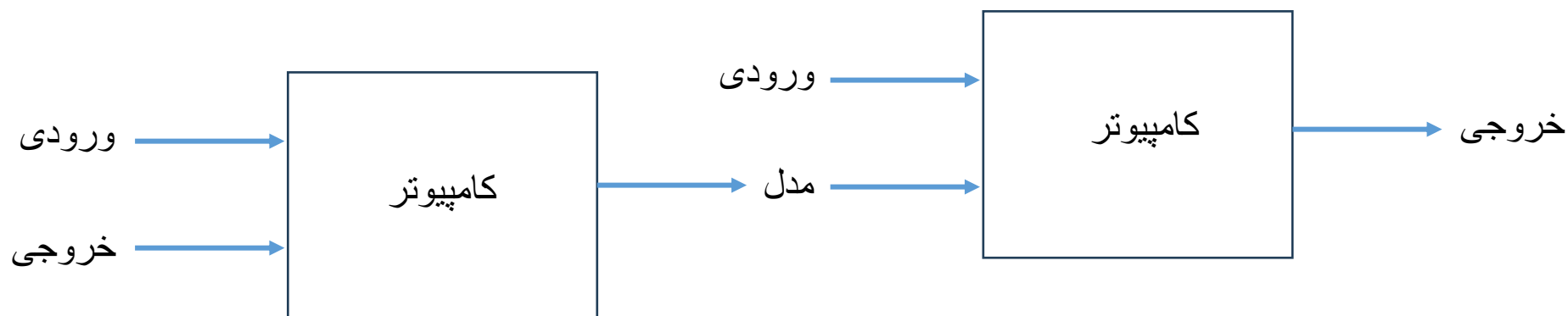
1. یادگیری تقویتی (Reinforcement Learning)
2. یادگیری بانظارت (Supervised Learning)
3. یادگیری بدون نظارت (Unsupervised Learning)
4. یادگیری نیمه نظارتی (Semi-Supervised Learning)
5. یادگیری خودنظارتی (Self-Supervised Learning)

# یادگیری ماشین

□ **یادگیری ماشین** شاخه‌ای از هوش مصنوعی است که به کامپیوترها امکان می‌دهد با استفاده از داده‌ها، الگوهایی را بیاموزند و از این آموخته‌ها برای تحلیل یا پیش‌بینی روی داده‌های جدید استفاده کنند.

## □ مثال:

- ورودی ← داده‌های ورودی (مثلاً عکس)
- خروجی ← برچسب درست (مثلاً گربه، سگ و ..)
- مدل ← تابعی که نگاشت ورودی به خروجی را یاد گرفته است.



# یادگیری تقویتی (Reinforcement Learning)

□ عامل با جمع‌آوری تجربه‌هایی شامل حالت، عمل، پاداش و حالت بعد، مدلی می‌آموزد که بتواند از طریق آن، در هر حالت (وضعیت) بهترین عمل را انتخاب کند.

□ هدف: انجام اعمالی برای بیشینه‌سازی پاداش کلی



□ معروف‌ترین الگوریتم‌های یادگیری تقویتی

▪ یادگیری Q (Q-Learning)