

رسالة محمد



یادگیری عمیق

مدرس: محمدرضا محمدی

بهار ۱۴۰۲

روش‌شناسی کاربردی

Practical Methodology

انتخاب ابرپارامترها

- در طراحی و آموزش هر شبکه ابرپارامترهای بسیار زیادی از جمله نرخ آموزش، تعداد لایه‌ها، تعداد واحدها، توابع فعال‌سازی، تابع ضرر، ابعاد فیلترها و ... وجود دارند

- تاثیر بر زمان و حافظه اجرای الگوریتم

- تاثیر بر کیفیت مدل آموزش‌دیده

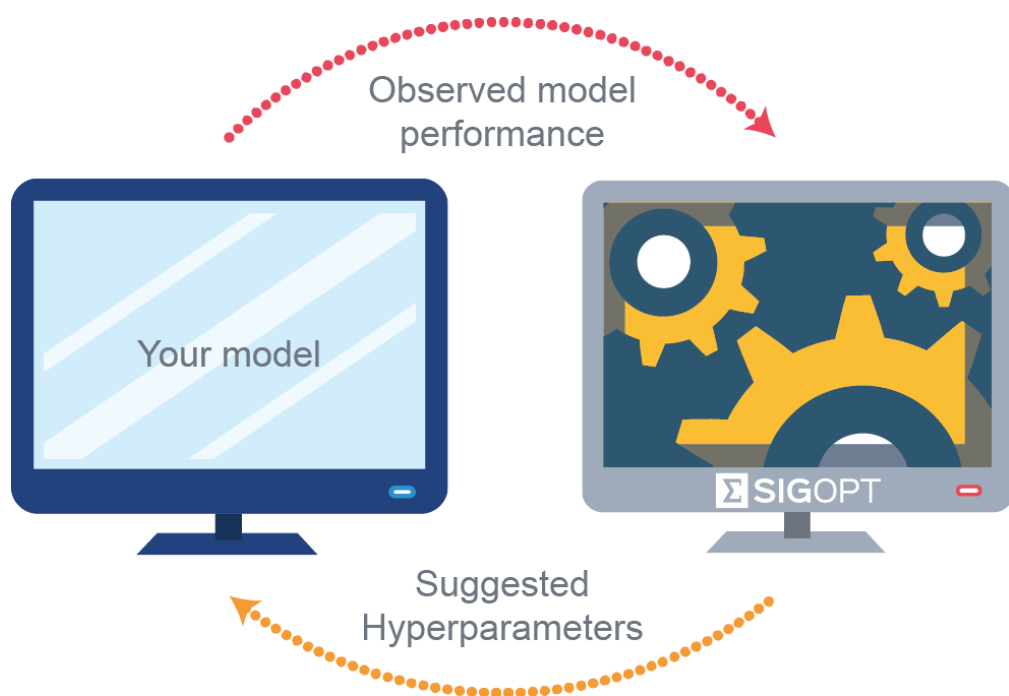
- دو رویکرد اصلی برای انتخاب ابرپارامترها وجود دارد:

- انتخاب دستی

- نیاز به درک ابرپارامترها دارد

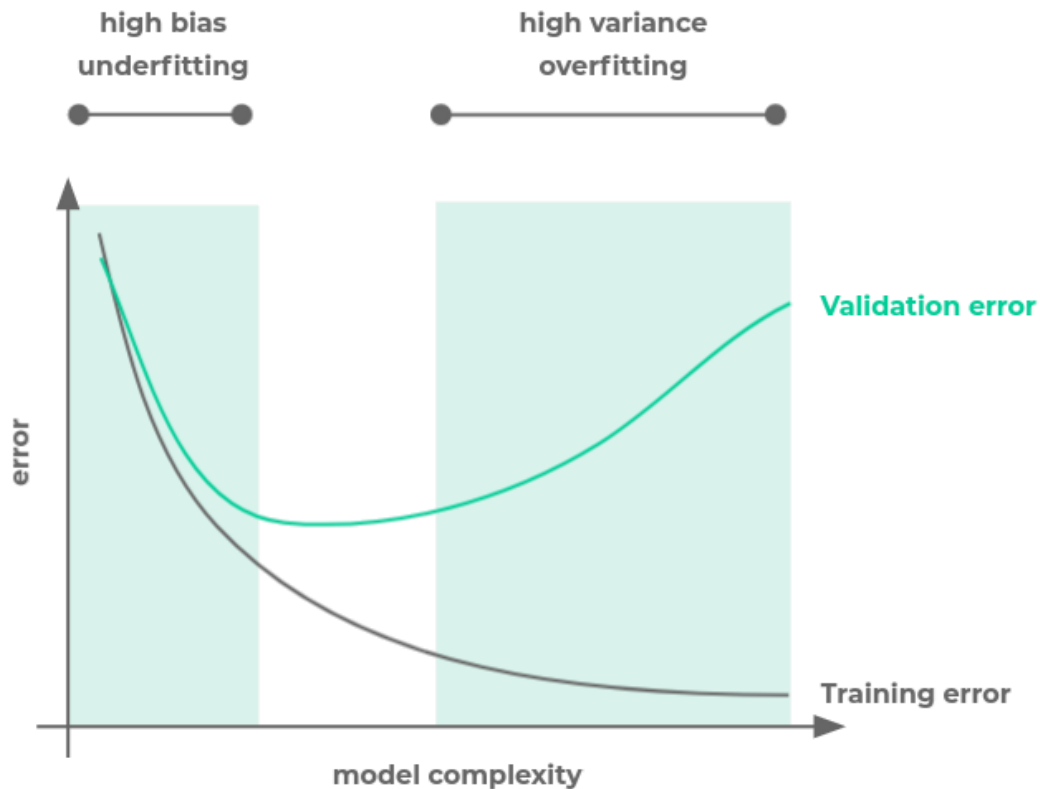
- انتخاب خودکار

- اغلب از نظر محاسباتی بسیار پرهزینه هستند



تنظیم دستی ابرپارامترها

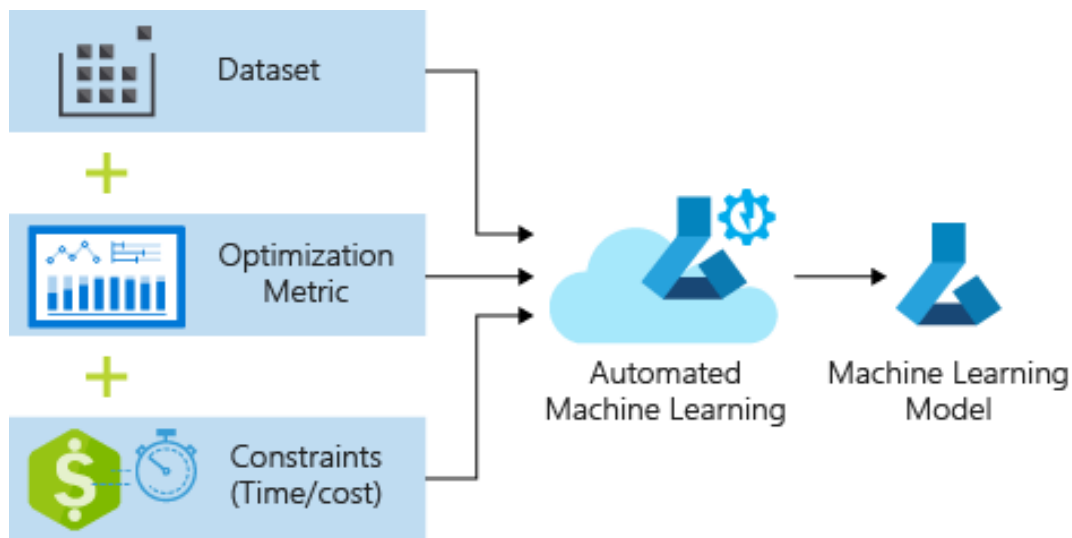
- نیاز است تا رابطه میان ابرپارامترها با خطای آموزش، خطای تعمیم‌دهی، و منابع محاسباتی (زمان و حافظه) درک شده باشد



ابریارمتر	ظرفیت مدل افزایش می‌یابد اگر ...
تعداد لایه‌های میانی	
ضریب کاهش وزن	
ابعاد کرنل کانولوشنی	
نرخ Dropout	
نرخ آموزش	

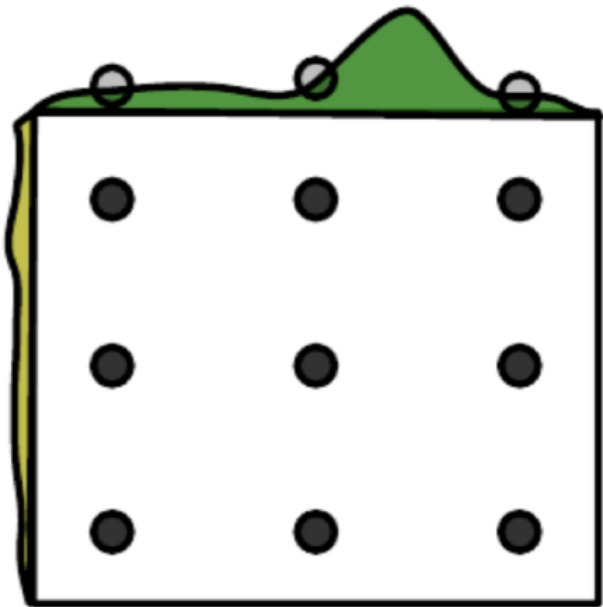
تنظیم خودکار ابرپارامترها

- تنظیم دستی ابرپارامترها می‌تواند بسیار خوب کار کند اگر کاربر دارای تجربه کافی باشد و نقطه شروع خوبی داشته باشد
- الگوریتم یادگیری ماشین ایده‌آل فقط یک مجموعه داده و اهداف را می‌گیرد و تابع تبدیل را بدون نیاز به تنظیم دستی ابرپارامترها محاسبه می‌کند



جستجوی شبکه‌ای

- اگر تعداد ابرپارامترها سه یا کمتر باشد
- برای هر ابرپارامتر، مجموعه محدودی از مقادیر برای جستجو انتخاب می‌شود
 - مانند $lr \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$
- هزینه محاسباتی بالا
 - $O(n^m)$
- معمولاً زمانی بهترین عملکرد را دارد که در چندین مرحله انجام شود
 - در هر مرحله محدوده جستجو کوچک‌تر و دقیق‌تر می‌شود



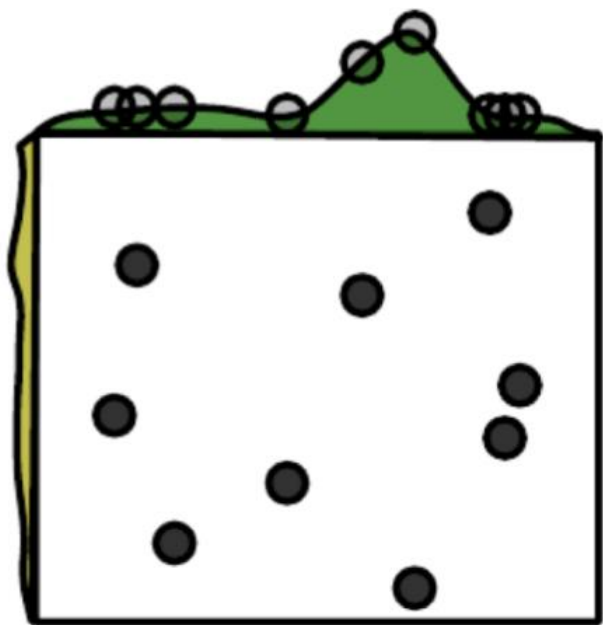
جستجوی تصادفی

- پیاده‌سازی ساده، و همگرایی سریع‌تر
- برای هر ابرپارامتر یک توزیع تصادفی تعریف می‌شود

$$\log(lr) \sim u(-1, -5) -$$

$$lr = 10^{\log(lr)} -$$

- مجموعه بزرگتری از مقادیر جستجو می‌شود
- اگر برخی ابرپارامترها اثر زیادی نداشته باشند، بسیار کارآمدتر است



بهینه‌سازی ابرپارامتر مبتنی بر مدل

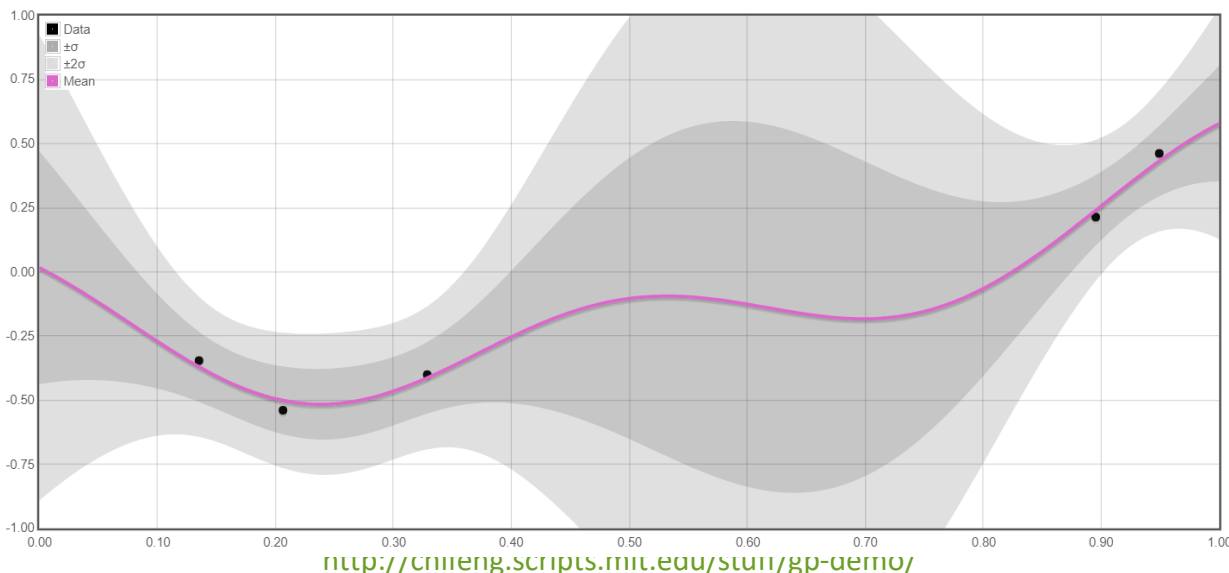
- جستجوی ابرپارامترهای مناسب را می‌توان به عنوان یک مسئله بهینه‌سازی در نظر گرفت
 - متغیرهای بهینه‌سازی، ابرپارامترها هستند
 - تابعی که باید بهینه شود خطای مجموعه اعتبارسنجی است
 - متأسفانه، در اکثر موارد گرادیان در دسترس نیست

$$A^* = \arg \min_{A \in \mathcal{A}} \mathcal{L}_{val}(A(w^*), D_{val})$$

$$s. t. \quad w^* = \arg \min_w \mathcal{L}_{train}(A(w), D_{train})$$

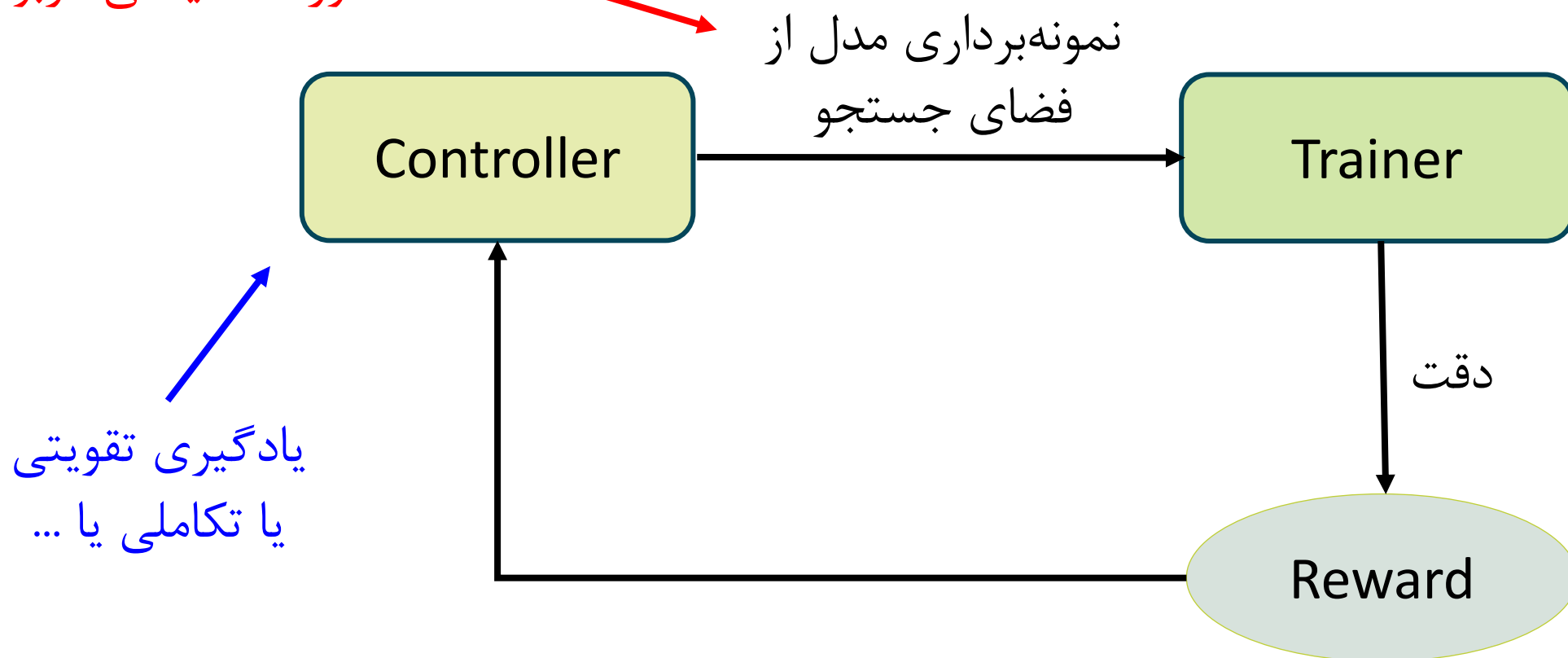
بهینه‌سازی ابرپارامتر مبتنی بر مدل

- می‌توان بر اساس آزمایش‌های انجام شده، مدلی برای خطای مجموعه اعتبارسنجی ایجاد کرد
 - با انجام بهینه‌سازی در این مدل، ابرپارامترهای جدیدی پیشنهاد خواهند شد
- می‌توان از یک مدل رگرسیون Bayesian برای تخمین مقدار مورد انتظار و همچنین عدم قطعیت آن به ازای هر ابرپارامتر استفاده کرد
- برای این بهینه‌سازی باید تعادلی ایجاد شود میان:
 - اکتشاف (Exploration)
 - بهره‌برداری (Exploitation)



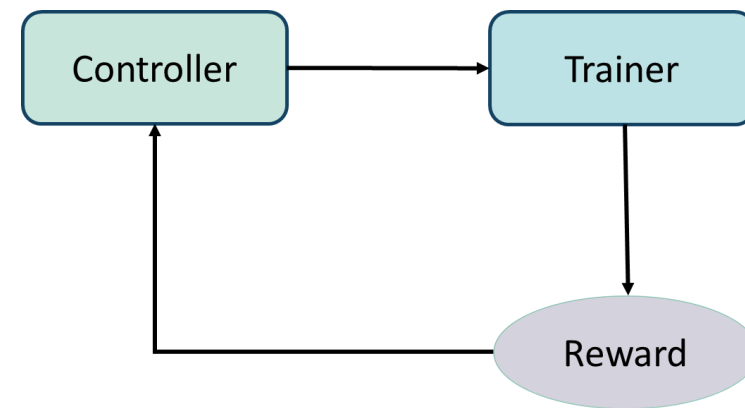
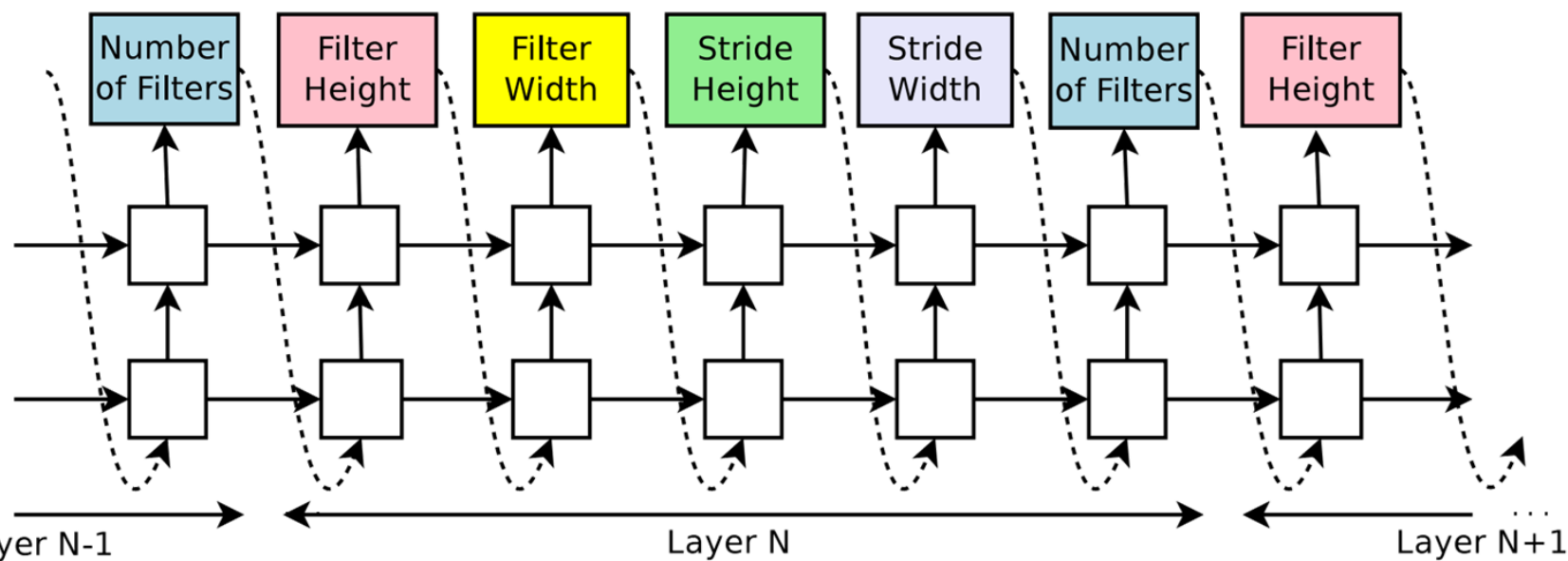
جستجوی معماری عصبی (Neural Architecture Search)

بر اساس اصول اولیه در
حوزه تحقیقاتی مربوطه



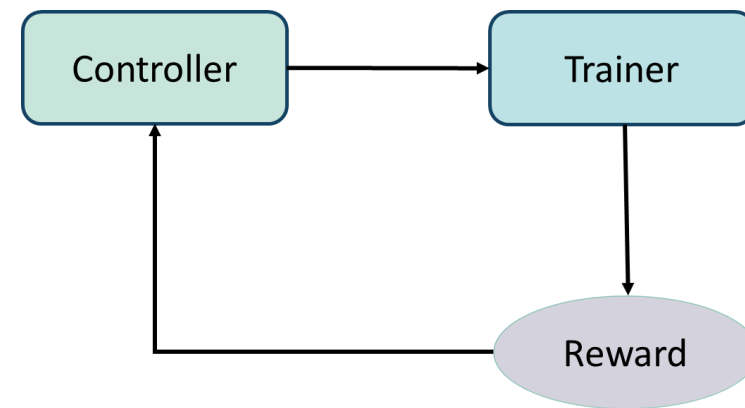
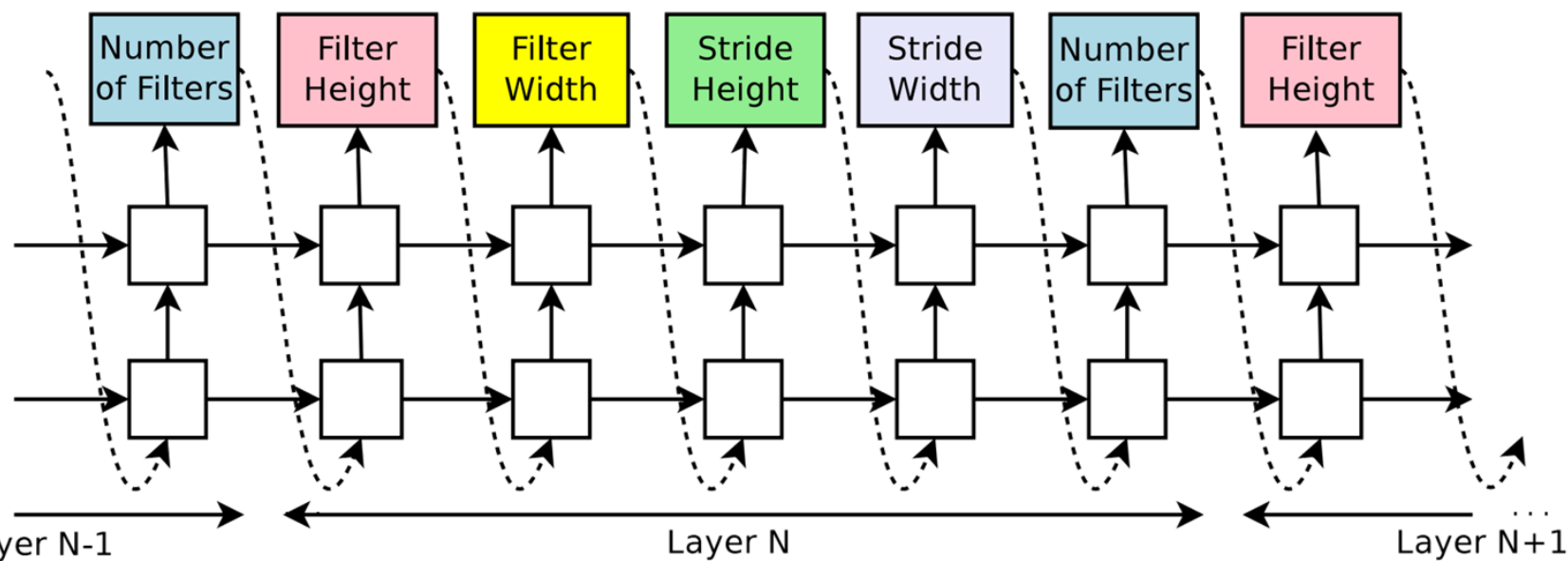
جستجوی معماری عصبی (Neural Architecture Search)

- در کنترلر، از یک شبکه بازگشتی برای تولید ابرپارامترهای شبکه کانولوشنی استفاده می‌شود
- تعداد لایه‌های شبکه در هر مرحله ثابت فرض می‌شود که در طول زمان افزایش می‌یابد
- پس از همگرایی شبکه کانولوشنی، دقت آن بر روی مجموعه اعتبارسنجی محاسبه شده و پارامترهای کنترلر به روز می‌شوند



جستجوی معماری عصبی (Neural Architecture Search)

- برای به روزرسانی پارامترهای کنترلر از یادگیری تقویتی استفاده می‌شود تا دقت اعتبارسنجی مورد انتظار را بیشینه کند



جستجوی مبتنی بر سلول

- در بسیاری از روش‌های موجود، یک ساختار پایه (سلول) در طراحی شبکه استفاده می‌شود
- بهینه‌سازی برای سلول انجام می‌شود

