

# اهمیت ویژگی‌ها در یادگیری ماشین

سجاد سبزی  
محمدرضا احمدی تشنیزی

۱۰ آذر ۱۴۰۲

## ۱ تعریف و اهمیت

### ۱.۱ تعریف

اهمیت ویژگی‌ها میزان مشارکت هر ویژگی در پیش‌بینی یک مدل یادگیری ماشین را اندازه‌گیری می‌کند. این به درک داده‌ها، مدل و ارتباط بین آنها کمک می‌کند.

### ۲.۱ اهمیت در یادگیری ماشین

دانستن کدام ویژگی‌ها بیشترین تأثیر را بر روی مدل دارند می‌تواند در بهبود عملکرد مدل، بهتر شناختن داده‌ها و افزایش قابلیت تفسیر مدل کمک کند.

## ۲ تکنیک‌ها برای ارزیابی اهمیت ویژگی‌ها

- روش‌های مبتنی بر مدل: برخی الگوریتم‌ها به طور ذاتی اهمیت ویژگی‌ها را به عنوان بخشی از خروجی مدل ارائه می‌دهند. مثال‌ها شامل درخت‌های تصمیم، جنگل‌های تصادفی و ماشین‌های گرادیانی می‌شود.
- اهمیت جایگشت: این روش شامل تصادفی کردن ترتیب ویژگی‌ها و اندازه‌گیری تغییر در عملکرد مدل است. تغییرات معنادار به معنی اهمیت بالا هستند.
- SHAP SHapley Additive exPlanations: رویکردی بازی‌نظری برای توضیح خروجی هر مدل یادگیری ماشین با اختصاص مقدار اهمیت به هر ویژگی.
- LIME Local Interpretable Model-agnostic Explanations: توضیح پیش‌بینی‌های هر طبقه‌بند در یک روش قابل تفسیر و وفادار با تقریب آن به صورت محلی با یک مدل قابل تفسیر.

## ۳ عوامل موثر بر اهمیت ویژگی‌ها

- کیفیت داده: نویز و خطاها در داده‌ها می‌توانند بر اهمیت ویژگی‌ها تأثیر بگذارند.
- پیچیدگی مدل: مدل‌های مختلف ممکن است برای همان مجموعه ویژگی‌ها ترتیب‌های مختلفی از اهمیت ارائه دهند.
- همبستگی: ویژگی‌های با همبستگی بالا می‌توانند اندازه‌گیری‌های اهمیت را اشتباه کنند.

## ۴ کاربردها

- ساده‌سازی مدل: حذف ویژگی‌های کم اهمیت می‌تواند مدل را ساده‌تر کند بدون کاهش قابل توجه در عملکرد.
- درک عوامل تأثیرگذار: در حوزه‌هایی مانند پزشکی یا مالی، درک ویژگی‌های تأثیرگذار می‌تواند برای تصمیم‌گیری‌ها بسیار حیاتی باشد.
- مهندسی ویژگی: شناسایی ویژگی‌های مهم می‌تواند در ایجاد ویژگی‌های جدید که عملکرد مدل را بهبود بخشند، راهنمایی کند.

## ۵ چالش‌ها و ملاحظات

- قابل تفسیری در مقابل دقت: برخی مدل‌های بسیار دقیق مانند یادگیری عمیق ممکن است اهمیت ویژگی‌ها را به صورت واضح ارائه ندهند.
- تعصب و انصاف: معیارهای اهمیت ممکن است تعصب داشته باشند اگر داده‌های آموزش نمایانگر جامعه نباشند.
- وابستگی به نوع مدل: اهمیت ویژگی‌ها وابسته به نوع مدل است، به این معنا که مدل‌های مختلف ممکن است به ویژگی‌های یکسان اهمیت‌دهی متفاوتی داشته باشند.

## ۶ روش‌های بهتر

- اعتبارسنجی متقابل: از اعتبارسنجی متقابل استفاده کنید تا اهمیت ویژگی‌ها را ارزیابی کنید و از بیش‌برازش جلوگیری کنید.
- تحلیل مقایسه‌ای: اهمیت ویژگی‌ها را در مدل‌های مختلف مقایسه کنید تا دیدگاه جامع‌تری داشته باشید.
- دانش حوزه‌ای را مد نظر قرار دهید: دانش حوزه‌ای را به منظور تفسیر اهمیت ویژگی‌ها به صورت صحیح به مد نظر قرار دهید.

## ۷ جهت‌های آینده

- تحقیق: تحقیقات در زمینه هوش تفسیری (XAI) به طور مداوم در حال بهبود روش‌های درک و تفسیر اهمیت ویژگی‌ها هستند.
- استانداردسازی: تلاش‌هایی برای استانداردسازی معیارهای اهمیت ویژگی برای بهترین مقایسه و قابلیت تکرار انجام می‌شود.

## ۸ مخازن اصلی اهمیت ویژگی در PyTorch

در این مستند، چند مخزن GitHub مرتبط با اهمیت ویژگی و تفسیرپذیری در PyTorch فهرست شده است.

۱. `PyTorch/Captum`: یک کتابخانه جامع برای تفسیر مدل در PyTorch است که ابزارهایی برای اهمیت ویژگی و تفسیرپذیری مدل ارائه می‌دهد. این کتابخانه در جامعه PyTorch خوب پشتیبانی می‌شود و محبوب است. مخزن GitHub
۲. `EthicalML/XAI`: این مخزن، ابزارهایی برای هوش مصنوعی قابل تفسیر را شامل می‌شود که به تفسیر اهمیت ویژگی نزدیک می‌شود. این مخزن بر روی مسائل اخلاقی و شفافیت در مدل‌های یادگیری ماشین تمرکز دارد. مخزن GitHub