

دانشگاه ملی مهارت

آموزشکده میناب

تمرینات بخش اول

نام و نام خانوادگی: هدا مهرانی پور_مرضیه نجفی

واحد درسی : مباحث ویژه

رشته : مهندسی کامپیوتر

مدرس : محمد احمد زاده

بهمن ۱۴۰۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

A. Supervised Learning و Unsupervised Learning چه تفاوتی دارند؟

در یادگیری نظارت‌شده، الگوریتم با استفاده از مجموعه داده‌های برچسب‌دار (دارای ورودی و خروجی مشخص) آموزش می‌بیند تا تابعی برای پیش‌بینی یا دسته‌بندی یاد بگیرد، مانند رگرسیون یا طبقه‌بندی. در مقابل، یادگیری بدون نظارت با داده‌های بدون برچسب کار می‌کند و هدفش استخراج الگوها یا ساختارهای ذاتی، مانند خوشه‌بندی یا کاهش ابعاد، بدون هدایت مستقیم است.

B. چرا Feature Scaling در الگوریتم‌های Machine Learning ضروری است؟

Feature Scaling در الگوریتم‌های یادگیری ماشین ضروری است زیرا مقیاس‌های متفاوت ویژگی‌ها می‌توانند تأثیر نامتوازنی بر بهینه‌سازی (مثل گرادیان نزولی) بگذارند و همگرایی را کند یا ناپایدار کنند. با استانداردسازی یا نرمال‌سازی، ویژگی‌ها در محدوده یکسان قرار می‌گیرند و عملکرد الگوریتم‌هایی مثل SVM یا KNN که به فاصله حساس‌اند، بهبود می‌یابد.

C. Standardization و Normalization چه تفاوتی دارند؟

Standardization مقادیر ویژگی‌ها را به گونه‌ای تغییر می‌دهد که میانگین صفر و انحراف معیار یک داشته باشند (تبدیل به توزیع نرمال استاندارد)، مناسب برای الگوریتم‌های حساس به مقیاس مثل گرادیان نزولی. Normalization مقادیر را به یک بازه مشخص (معمولاً $[0, 1]$) می‌برد، که برای الگوریتم‌های مبتنی بر فاصله مثل KNN یا شبکه‌های عصبی که به مقادیر محدود نیاز دارند، ایده‌آل است.

D. چرا Min-Max Normalization برای مقیاس‌بندی داده‌ها استفاده می‌شود؟

Min-Max Normalization داده‌ها را به بازه‌ای مشخص (معمولاً $[0, 1]$) مقیاس‌بندی می‌کند تا اختلافات مقیاس بین ویژگی‌ها حذف شود و همه در یک محدوده یکسان قرار گیرند. این روش برای الگوریتم‌هایی مثل شبکه‌های عصبی یا KNN که به مقادیر محدود و یکنواخت حساس‌اند، ضروری است تا از تسلط ویژگی‌های با مقیاس بزرگ جلوگیری کند.

E. Z-Score Normalization چیست و چرا کاربرد دارد؟

Z-Score Normalization یک روش استانداردسازی داده‌ها است که در آن هر مقدار داده با کم کردن میانگین و تقسیم بر انحراف معیار، به یک توزیع نرمال با میانگین صفر و انحراف معیار یک تبدیل می‌شود. این روش برای مقایسه داده‌هایی با واحدها یا محدوده‌های متفاوت، حذف سوگیری ناشی از مقیاس‌های مختلف و بهبود عملکرد الگوریتم‌های یادگیری ماشین (مانند KNN یا PCA) ضروری است، زیرا داده‌ها را در یک فضای یکنواخت و قابل تفسیر قرار می‌دهد.

F. Regularization در الگوریتم‌های Machine Learning چیست؟

Regularization تکنیکی در یادگیری ماشین است که با افزودن یک جریمه (مثل $L1$ یا $L2$) به تابع هزینه، پیچیدگی مدل را کنترل می‌کند تا از بیش‌برازش (Overfitting) جلوگیری شود. این روش وزن‌های مدل را محدود یا کوچک می‌کند، تعادل بین تطابق با داده‌های آموزشی و تعمیم‌پذیری به داده‌های جدید را بهبود می‌بخشد.

G. Overfitting و Underfitting چه مشکلاتی را در Model-building به وجود می‌آورند؟

Overfitting زمانی رخ می‌دهد که مدل بیش از حد به داده‌های آموزشی وابسته شود، جزئیات و نویز را یاد بگیرد و در داده‌های جدید عملکرد ضعیفی داشته باشد. Underfitting وقتی است که مدل بیش از حد ساده باشد، الگوهای اصلی داده‌ها را یاد نگیرد و در هر دو داده آموزشی و آزمایشی دقت پایینی داشته باشد.

H. Cross-Validation چرا در Train/Test Split کاربرد دارد؟

Cross-Validation در Train/Test Split استفاده می‌شود تا با تقسیم داده‌ها به چند زیرمجموعه و ارزیابی مدل در هر کدام، تخمین پایدارتر و قابل‌اعتمادتری از عملکرد مدل ارائه دهد. این روش از وابستگی نتیجه به یک تقسیم خاص جلوگیری می‌کند و به تشخیص Overfitting یا Underfitting کمک می‌کند.

۱. Gradient Descent چگونه کار می کند؟

Gradient Descent با محاسبه گرادیان (شیب) تابع هزینه نسبت به پارامترها، به صورت تکراری این پارامترها را در جهت کاهش خطا به روزرسانی می کند تا به کمینه محلی یا سراسری برسد. این روش از نرخ یادگیری (Learning Rate) برای کنترل اندازه گامها استفاده می کند تا تعادل بین سرعت و دقت بهینه سازی برقرار شود.

۲. چرا Deep Learning برای پیچیده ترین مسائل استفاده می شود؟

Deep Learning برای مسائل پیچیده استفاده می شود زیرا با لایه های متعدد عصبی، قادر است الگوهای سلسله مراتبی و غیرخطی عمیق را از داده های خام استخراج کند، مثل تشخیص تصویر یا زبان طبیعی. این توانایی در مدل سازی روابط پیچیده و خودکار سازی استخراج ویژگی ها، آن را از روش های سنتی متمایز می کند.