

Encoding Techniques (Label Encoding و One-Hot Encoding) چه تفاوتی دارند

ChatGPT گفت:

Label Encoding و One-Hot Encoding هر دو روش برای تبدیل داده های دسته های (داده های طبقه ای) به داده های عددی هستند، اما روش های انجام برای این کار هستند. تفاوت های اصلی آن ها عبارتند از:

رمزگذاری برچسب :

در این روش، به هر دسته (رده) یک عدد به فرد اختصاص داده می شود.

برای مثال، اگر یک ویژگی به نام «رنگ» داشته باشد که

این روش معمولاً استفاده می شود که ویژگی های دارای ارتباط ترتیبی باشد (مثل رتبه ها یا درجه ها).

یکی از معایب این روش این است که ممکن است مدلی باشد که ترتیب اعداد معنای خاصی دارد، حتی اگر اینطور نباشد.

رمزگذاری تک داغ :

در این روش، به هر دسته یک ستون جدید اختصاصی

برای مثال، اگر ویژگی «رنگ» با سه «قرمز»، «سبز» و «آبی» داشته باشد، یک ماتریس باینری به این صورت است:

کد را کپی کنید

قرمز سبز آبی

0 0 1

0 1 0

1 0 0

این روش در موردی که داده ها موارد خاصی ندارند (دسته های بدون ترتیب) بهتر عمل می کند.

معایب این روش شامل ایجاد تعداد ستون جدید است که ممکن است باعث افزایش مصرف حافظه و پیچیدگی محاسباتی شود.

تفاوت ها در خلاصه:

Label Encoding : تبدیل هر دسته به یک عدد مشخص به فرد. مناسب برای داده‌هایی که دارای مقادیر طبیعی هستند.

One-Hot Encoding : تبدیل هر دسته به یک ستون باینری خارجی. مناسب برای داده‌هایی که ترتیب نمی‌دهند.

در نهایت، انتخاب روش به نوع داده و نیاز مدل دارد.

ChatGPT ممکن است اشتباه کند. اطلاعات مهم را بررسی کنید

?