



الف) فرض کنید شبکه عصبی fully connected با یک hidden layer مانند شکل داریم.

در ابتدا هر یک از نرون ها دارای بردار w ای است که مؤلفه های آن به صورت random انتخاب شده اند.

در هر iteration ورودی ها در بردار w نرون ها ضرب داخلی می شوند و نتیجه از یک activation function عبور کرده و به عنوان ورودی به نرون های لایه بعدی داده می شود.

در انتهای iteration با مقایسه ی خروجی تخمین زده شدن و خروجی واقعی مقدار loss بدست می آید. حال با توجه به loss و ورودی باید وزن ها را طوری تغییر داد که مقدار loss کوچک و کوچکتر شود؛ که به این عمل backpropagation می گویند.

در انتهای iteration ما مقدار Error را داریم. برای کمتر کردن آن یکی از روش ها gradient descent می باشد که طی آن مقدار $\frac{\delta E}{\delta w_i}$ محاسبه می شود (تغییرات خط نسبت به وزن ها) آنها

و سپس این مقدار با یک ضریبی از w کم می شود تا مقدار خط با احتمال ورودی ها کمتر شود.

ب) مزایا → این روش به کمک تگ های مثل gradient-tape دارای performance خوبی است.
→ تعداد hyperparameter های این روش خیلی کم است و بیشتر به ورودی ها وابسته است.
→ کاملاً به صورت iterative و خودکار انجام می شود.

معایب → به داده های noise حساس است.
→ استفاده از gradient tape فضای حافظه زیادی می خواهد.
→ تعداد feature های بالا به طور چشمگیری performance را کاهش می دهد.
* این روش ممکن است در local minimum تابع مریض به شبکه گیر کند و وزن های بجینه و ایوآل بدست نیایند.