



شبکه‌های عصبی مصنوعی

جلسه هشتم:

پرسپترون چند لایه (۴)

(Multi-Layer Perceptron = MLP)

پرسپترون چندلایه (MLP)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.

پرسپترون چندلایه (MLP)

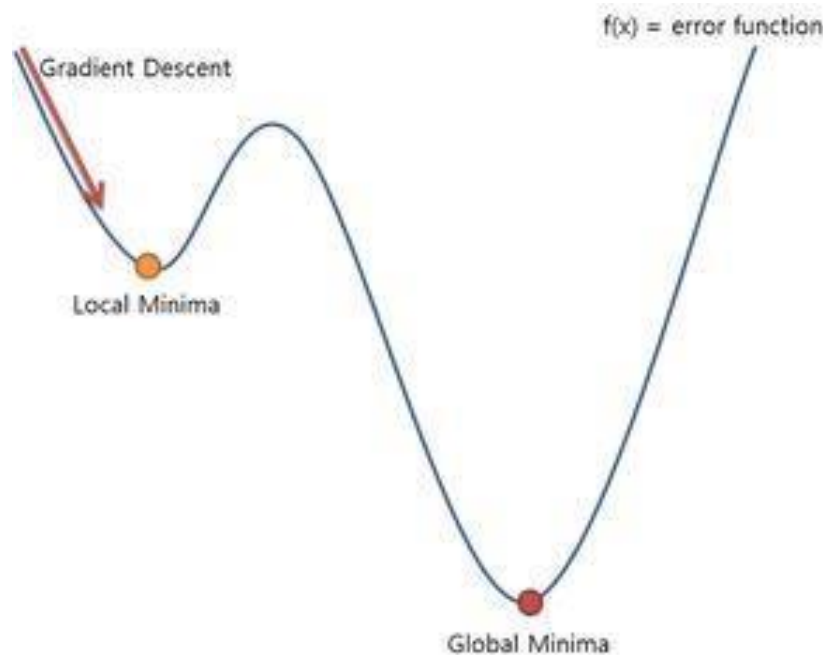
ضابطه توقف الگوریتم

- هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.
- فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

- هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.
- فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:
- ۱ - اندازه بردار گرادیان وزن ها $g(w^*)$ به مقداری کم رسیده باشد



پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.

– فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:

۱ – اندازه بردار گرادیان وزن ها $g(w^*)$ به مقداری کم رسیده باشد

$$\|g(w^*)\| \leq \varepsilon$$

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.

– فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:

۱ – اندازه بردار گرادیان وزن ها $g(w^*)$ به مقداری کم رسیده باشد

$$\|g(w^*)\| \leq \varepsilon$$

– معایب:

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.

– فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:

۱ – اندازه بردار گرادیان وزن ها $g(w^*)$ به مقداری کم رسیده باشد

$$\|g(w^*)\| \leq \epsilon$$

– معایب:

• احتمال دارد خیلی دیر به اندازه از پیش تعیین شده برسد.

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.

– فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:

۱ – اندازه بردار گرادیان وزن ها $g(w^*)$ به مقداری کم رسیده باشد

$$\|g(w^*)\| \leq \epsilon$$

– معایب:

- احتمال دارد خیلی دیر به اندازه از پیش تعیین شده برسد.
- نیاز به محاسبه بردار گرادیان دارد.

پرسپترون چندلایه (MLP)

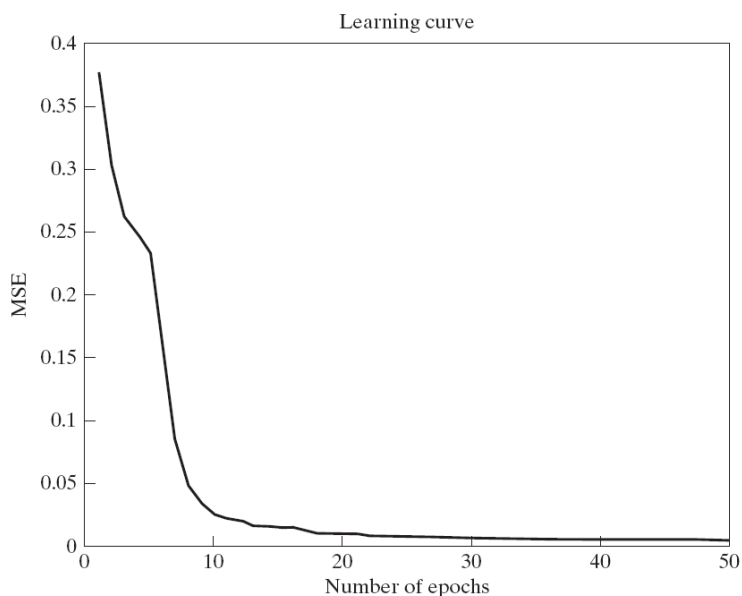
ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن‌ها همگرا شده‌اند، وجود ندارد. فقط می‌توان چند ضابطه برای این کار در نظر گرفت.

– فرض کنید \mathbf{w}^* بردار وزن‌های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس‌انتشار خطا متوقف شود هنگامی که:

۱ – اندازه بردار گرادیان وزن‌ها $g(\mathbf{w}^*)$ به مقداری کم رسیده باشد

۲ – میانگین مربعات خطا در یک دوره (epoch) \mathcal{E}_{av} به اندازه کافی کاهش یافته باشد.



پرسپترون چندلایه (MLP)

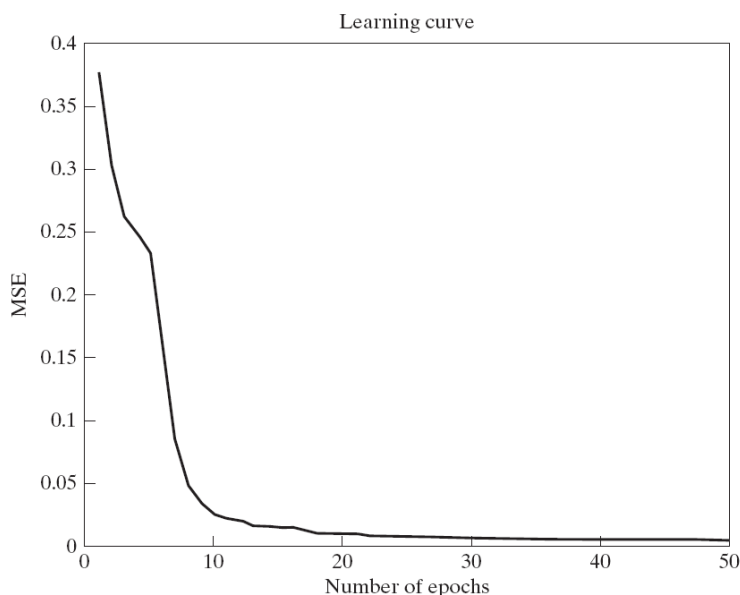
ضابطه توقف الگوریتم

– هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن‌ها همگرا شده‌اند، وجود ندارد. فقط می‌توان چند ضابطه برای این کار در نظر گرفت.

– فرض کنید \mathbf{w}^* بردار وزن‌های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس‌انتشار خطا متوقف شود هنگامی که:

۱ – اندازه بردار گرادیان وزن‌ها $g(\mathbf{w}^*)$ به مقداری کم رسیده باشد

۲ – میانگین مربعات خطا در یک دوره (epoch) \mathcal{E}_{av} به اندازه کافی کاهش یافته باشد.



عیب این ضابطه در این است که ممکن است هیچگاه به مقدار از پیش تعیین شده نرسد.

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

- هنوز هیچ قاعده مدونی برای این که مشخص شود چه وقت وزن ها همگرا شده اند، وجود ندارد. فقط می توان چند ضابطه برای این کار در نظر گرفت.
- فرض کنید w^* بردار وزن های بهینه شبکه برای حالتی که خطای شبکه به کمینه محلی یا جهانی رسیده باشد. در این صورت، الگوریتم پس انتشار خطا متوقف شود هنگامی که:
 - ۱ - اندازه بردار گرادیان وزن ها $g(w^*)$ به مقداری کم رسیده باشد
 - ۲ - میانگین مربعات خطا در یک دوره (epoch) \mathcal{E}_{av} به اندازه کافی کاهش یافته باشد.
 - ۳ - استفاده از هر دو ضابطه ۱ و ۲.

پرسپترون چندلایه (MLP)

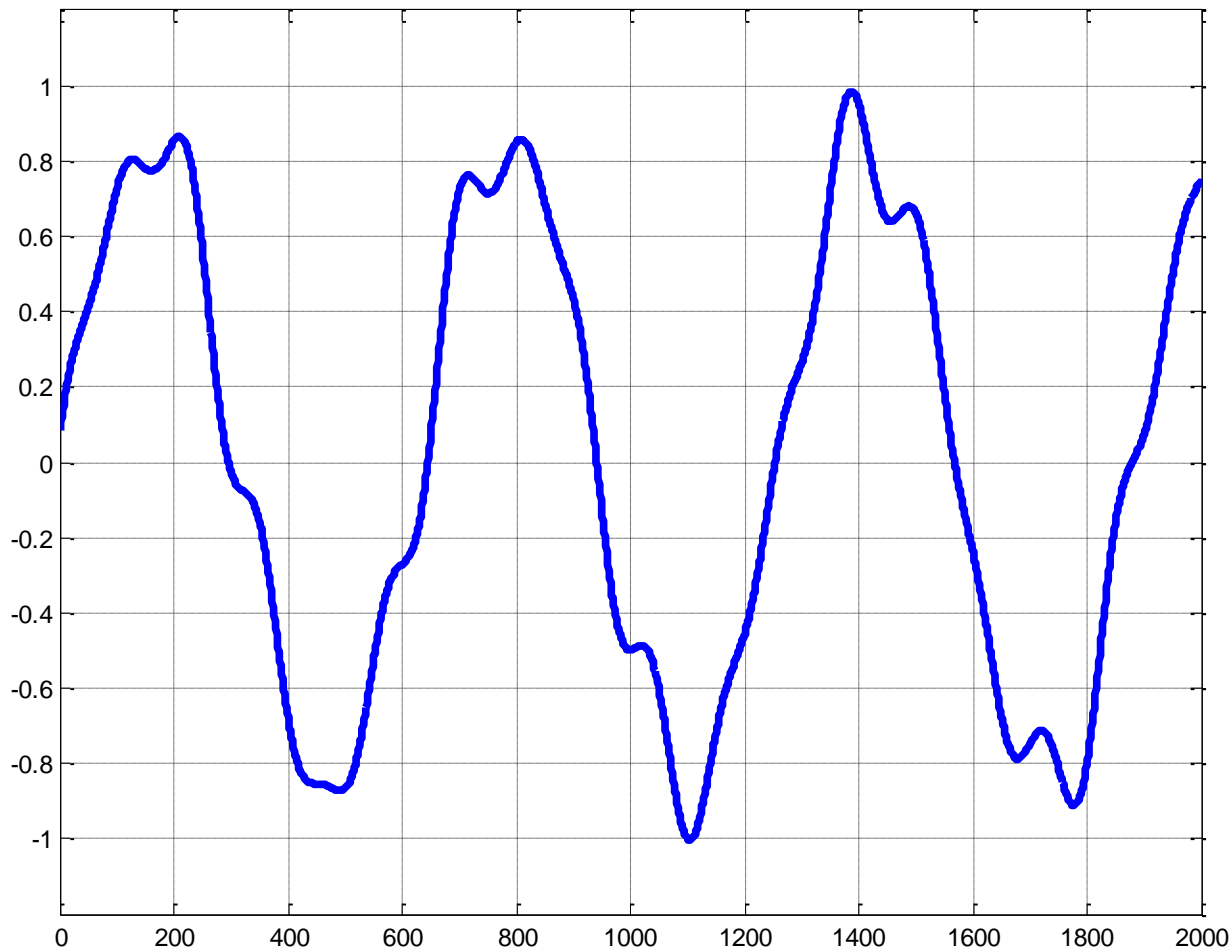
ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

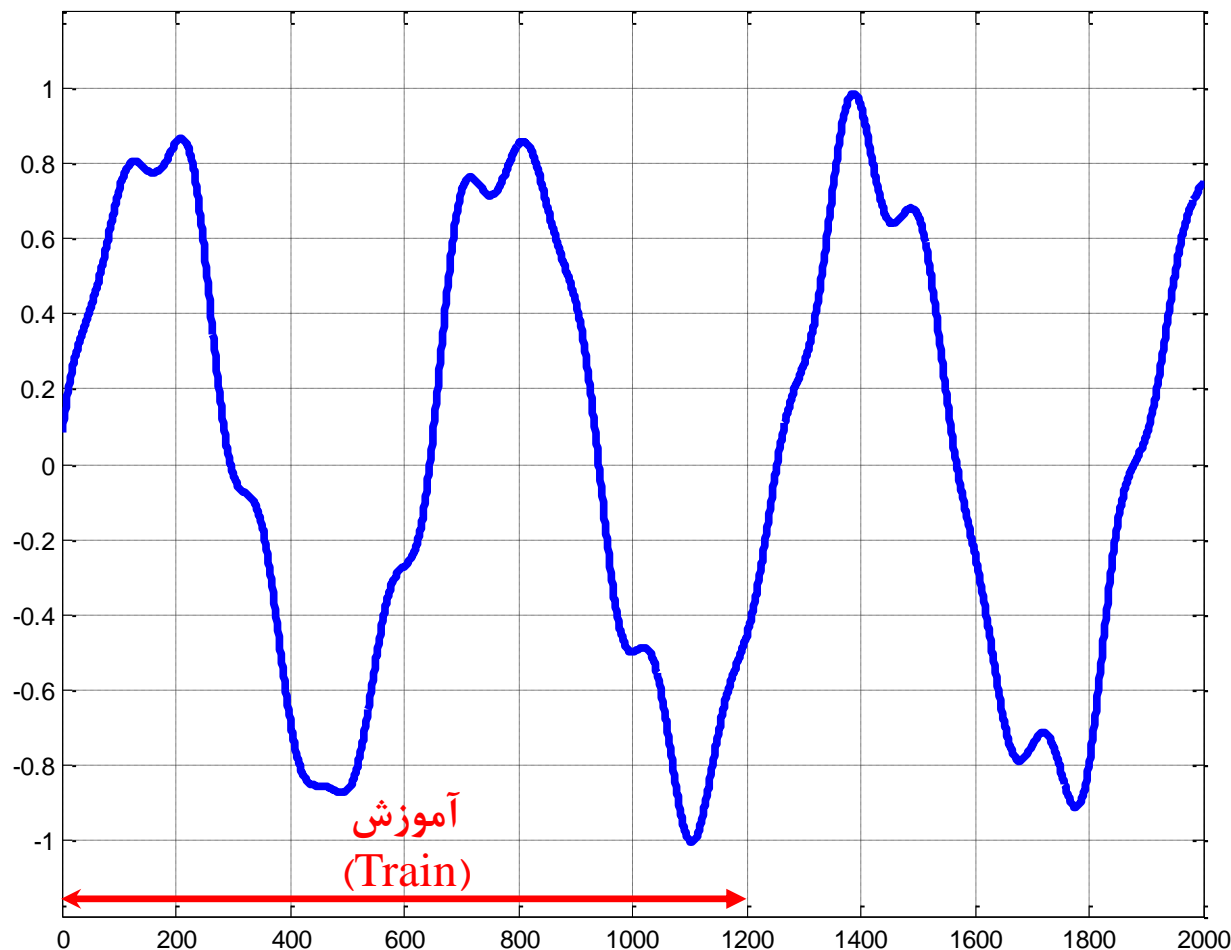


تقسیم داده‌ها به سه بخش:

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

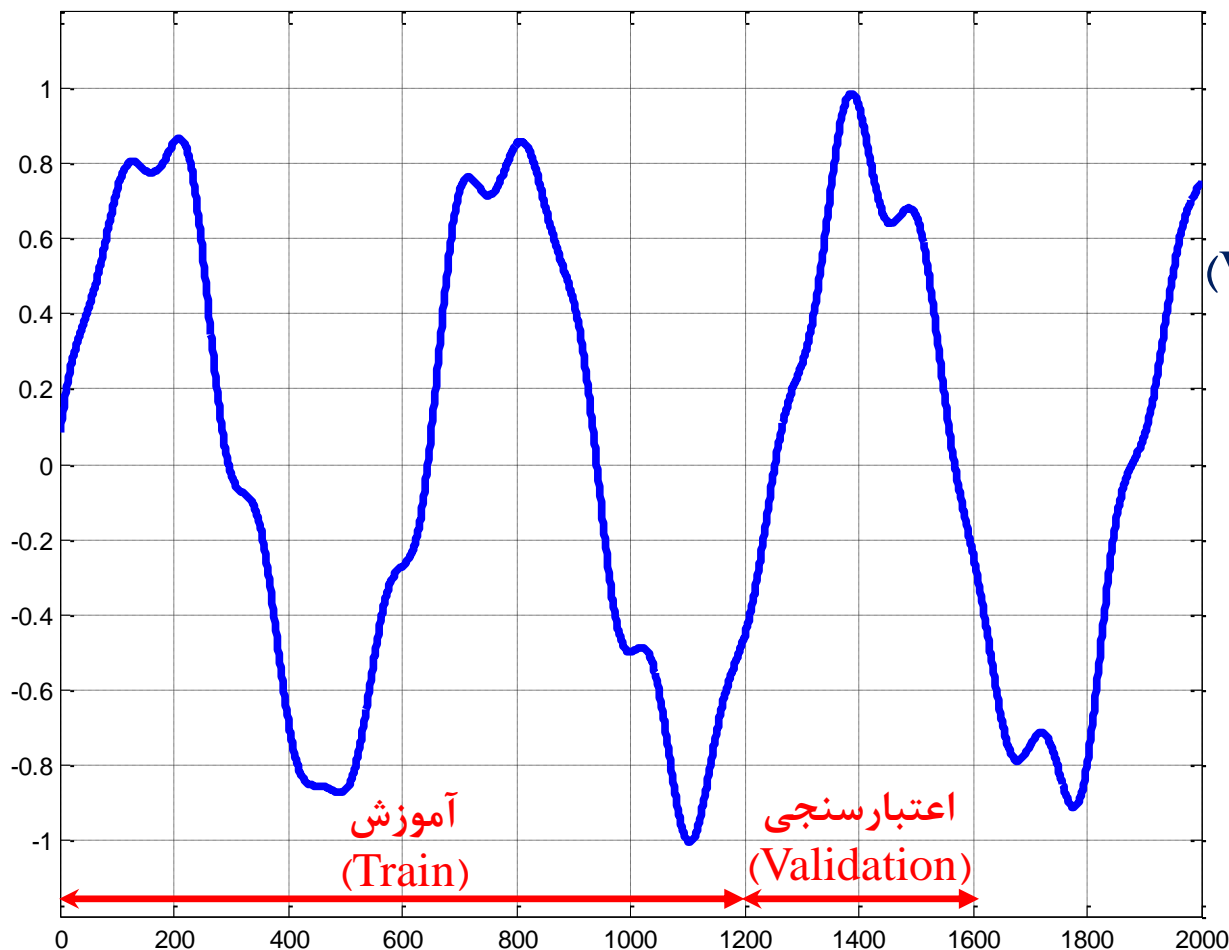


تقسیم داده‌ها به سه بخش:
• آموزش (Train)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)



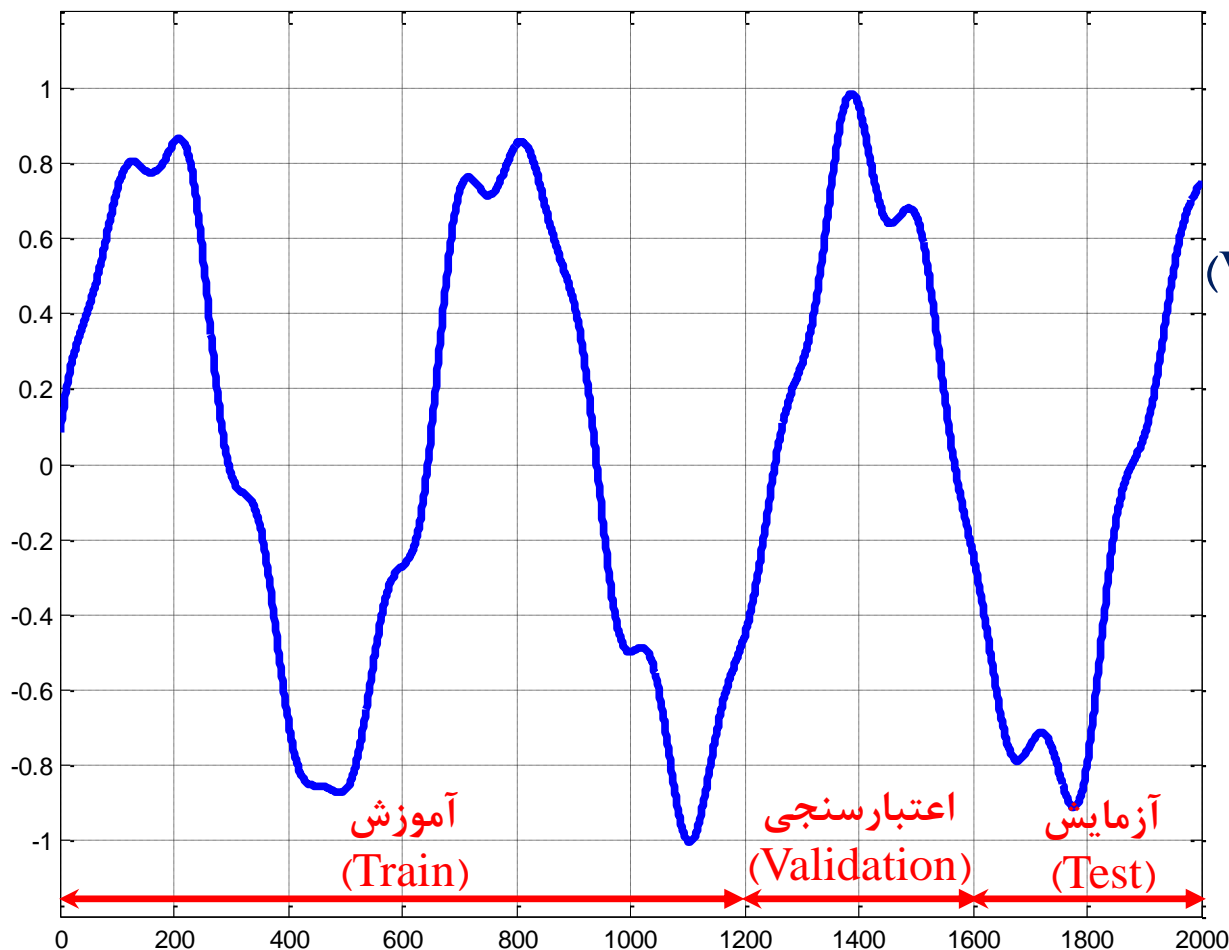
تقسیم داده‌ها به سه بخش:

- آموزش (Train)
- اعتبارسنجی (Validation)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)



تقسیم داده‌ها به سه بخش:

- آموزش (Train)
- اعتبارسنجی (Validation)
- آزمایش (Test)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

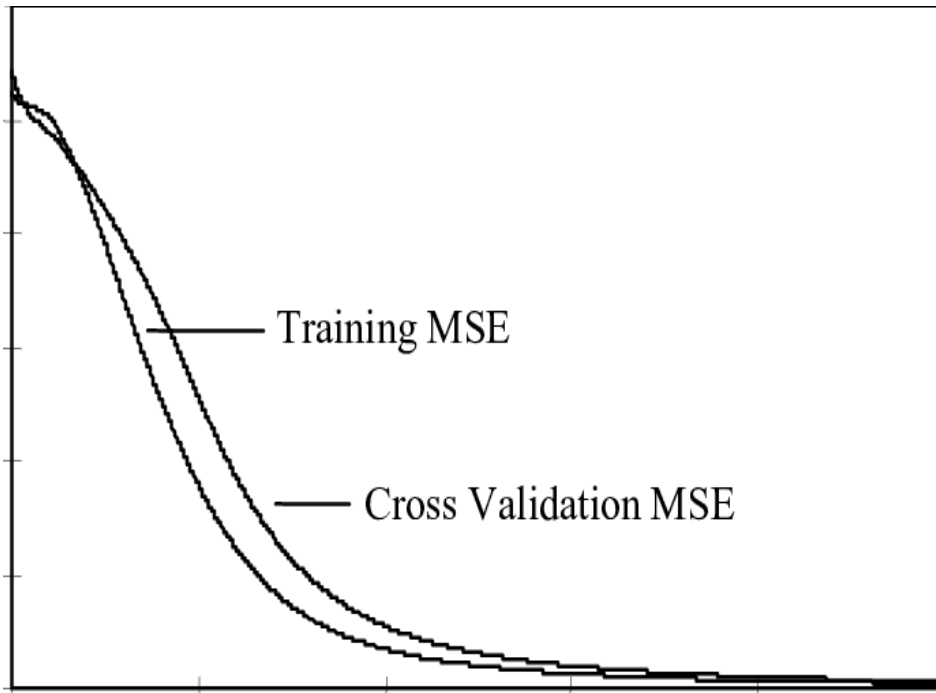
۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

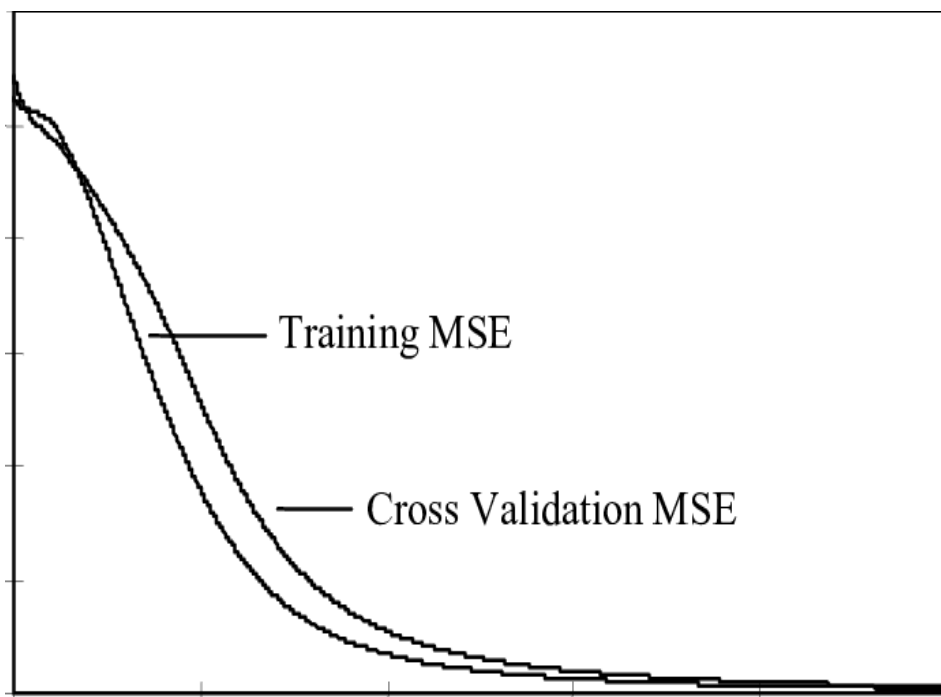
- به ازای هر چند دوره آموزش (مثلاً ۵ دوره)، وزن‌ها را ثابت نگه دارید و خطای مدل‌سازی شبکه را با داده‌های اعتبارسنجی به دست آورید.



پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

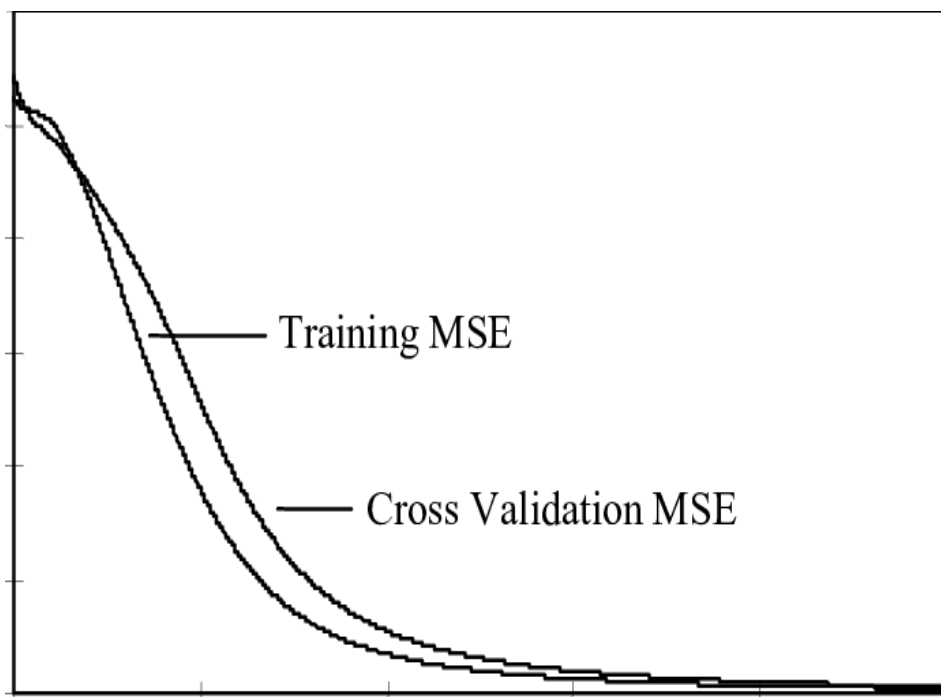


- به ازای هر چند دوره آموزش (مثلا ۵ دوره)، وزن‌ها را ثابت نگه دارید و خطای مدل‌سازی شبکه را با داده‌های اعتبارسنجی به دست آورید.
- آموزش شبکه هنگامی متوقف شود که خطای اعتبارسنجی به خطای آموزش نزدیک شود.

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

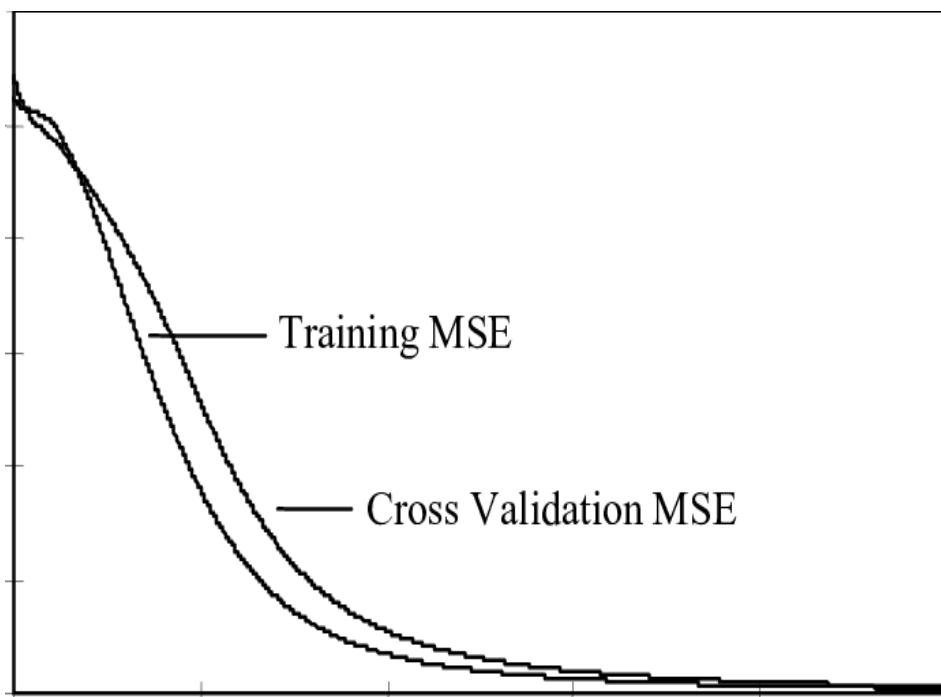


- به ازای هر چند دوره آموزش (مثلا ۵ دوره)، وزن‌ها را ثابت نگه دارید و خطای مدل‌سازی شبکه را با داده‌های اعتبارسنجی به دست آورید.
- آموزش شبکه هنگامی متوقف شود که خطای اعتبارسنجی به خطای آموزش نزدیک شود.
- در واقع در این جا باید مصالحه‌ای بین زمان آموزش و دقت تقریب در مدل‌سازی انجام داد.

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

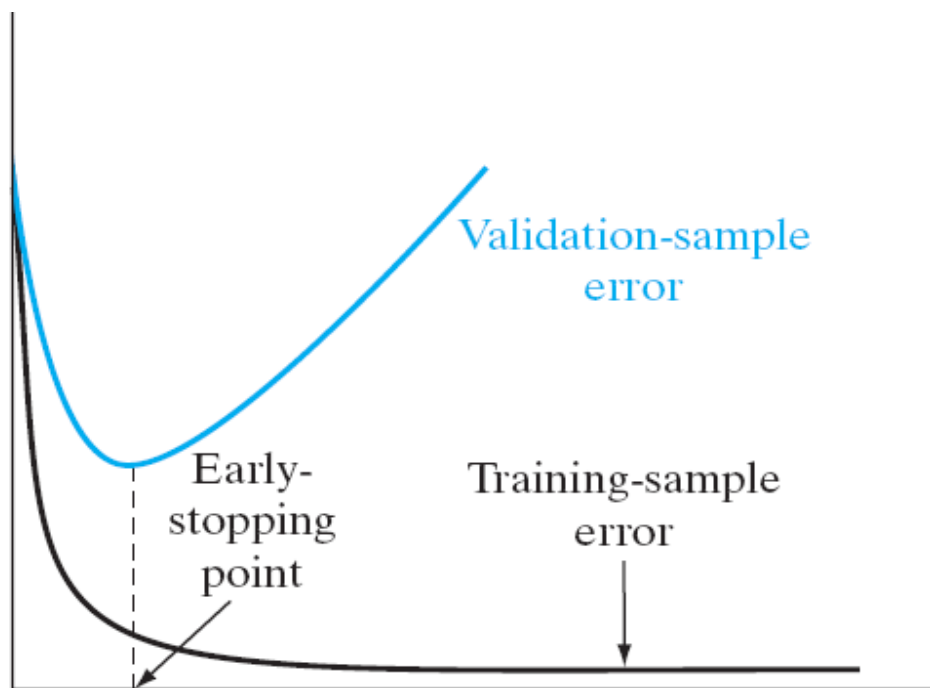


- به ازای هر چند دوره آموزش (مثلا ۵ دوره)، وزن‌ها را ثابت نگه دارید و خطای مدل‌سازی شبکه را با داده‌های اعتبارسنجی به دست آورید.
- آموزش شبکه هنگامی متوقف شود که خطای اعتبارسنجی به خطای آموزش نزدیک شود.
- در واقع در این جا باید مصالحه‌ای بین زمان آموزش و دقت تقریب در مدل‌سازی انجام داد.
- ولی باید توجه داشت که همواره نمودارهای آموزش و اعتبارسنجی به این شکل که در این جا نشان داده شده، نیست.

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)



- به ازای هر چند دوره آموزش (مثلاً ۵ دوره)، وزن‌ها را ثابت نگه دارید و خطای مدل‌سازی شبکه را با داده‌های اعتبارسنجی به دست آورید.

- آموزش شبکه هنگامی متوقف شود که خطای اعتبارسنجی به خطای آموزش نزدیک شود.

- در واقع در این جا باید مصالحه‌ای بین زمان آموزش و دقت تقریب در مدل‌سازی انجام داد.

- ولی باید توجه داشت که همواره نمودارهای آموزش و اعتبارسنجی به این شکل که در این جا نشان داده شده، نیست.

- یعنی پس از چند دوره، خطای اعتبارسنجی شروع به افزایش می‌کند و باید آموزش را زودتر از موعد متوقف کرد.

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

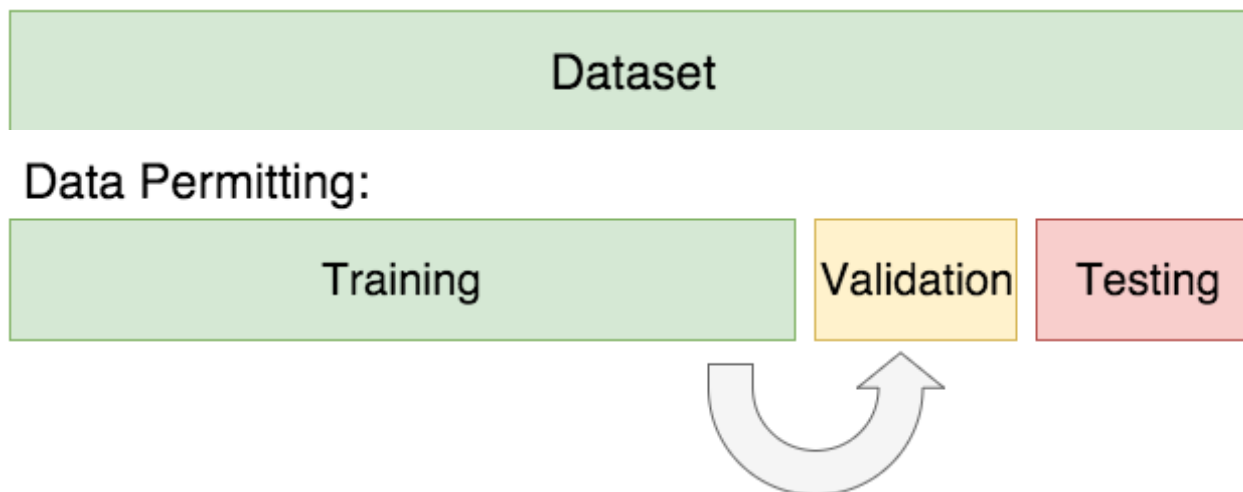
در مجموع، برای اعتبارسنجی باید چنین عمل کرد:

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

در مجموع، برای اعتبارسنجی باید چنین عمل کرد:

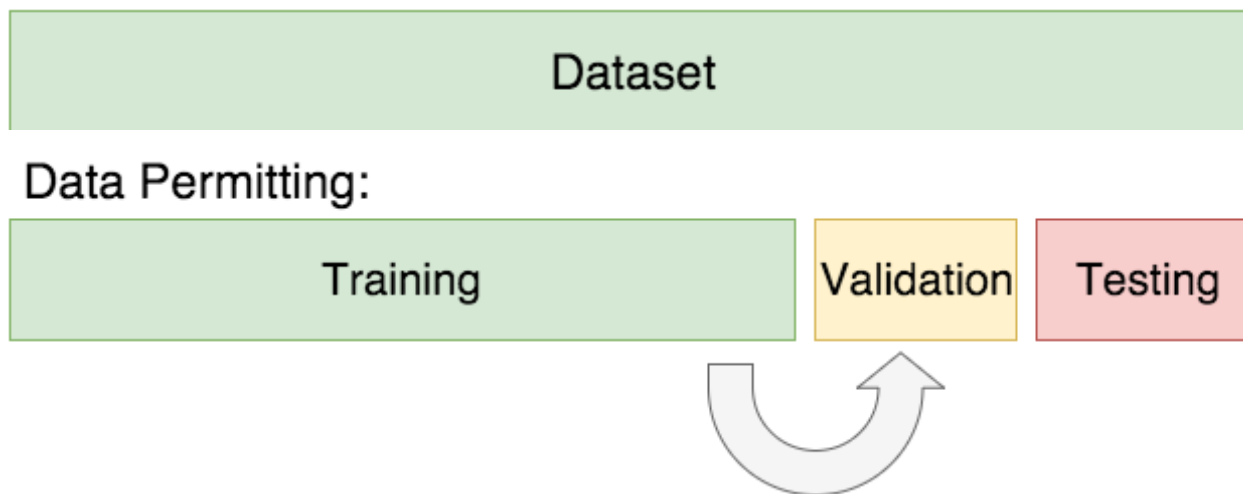


پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

در مجموع، برای اعتبارسنجی باید چنین عمل کرد:



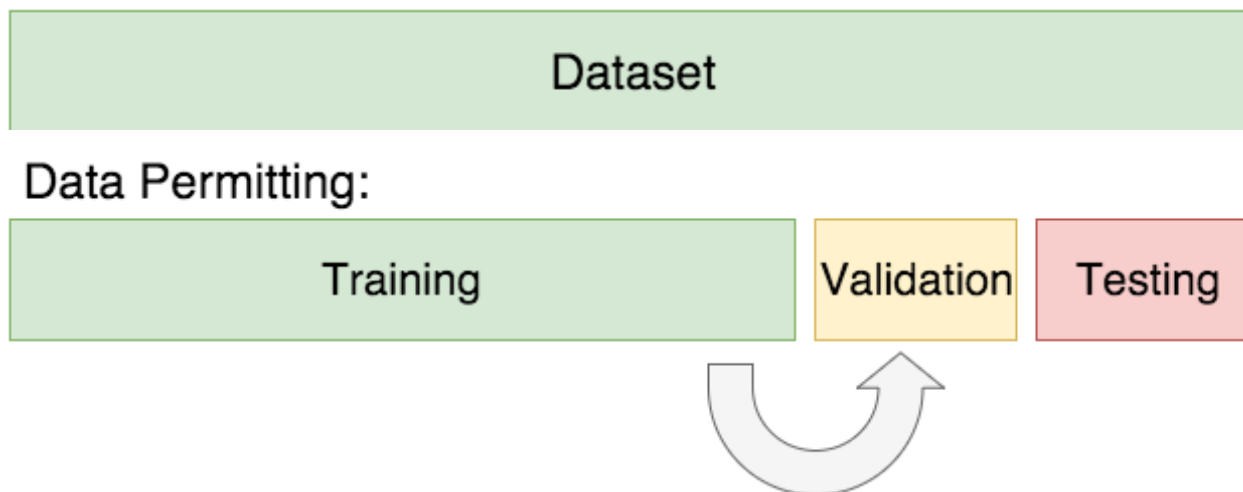
به این کار، اعتبارسنجی متقابل (Cross-Validation)

پرسپترون چندلایه (MLP)

ضابطه توقف الگوریتم

۴ - استفاده از داده‌های اعتبارسنجی متقابل (Cross-Validation)

در مجموع، برای اعتبارسنجی باید چنین عمل کرد:



به این کار، اعتبارسنجی متقابل (Cross-Validation)

نکته مهم:

همواره، داده‌های آموزش به صورت اتفاقی به شبکه اعمال شوند. البته این کار برای آموزش برخط امکان پذیر نیست.

پرسپترون چندلایه (MLP)

پرسپترون چندلایه (MLP)

غنی بودن داده‌های آموزش:

پرسپترون چندلایه (MLP)

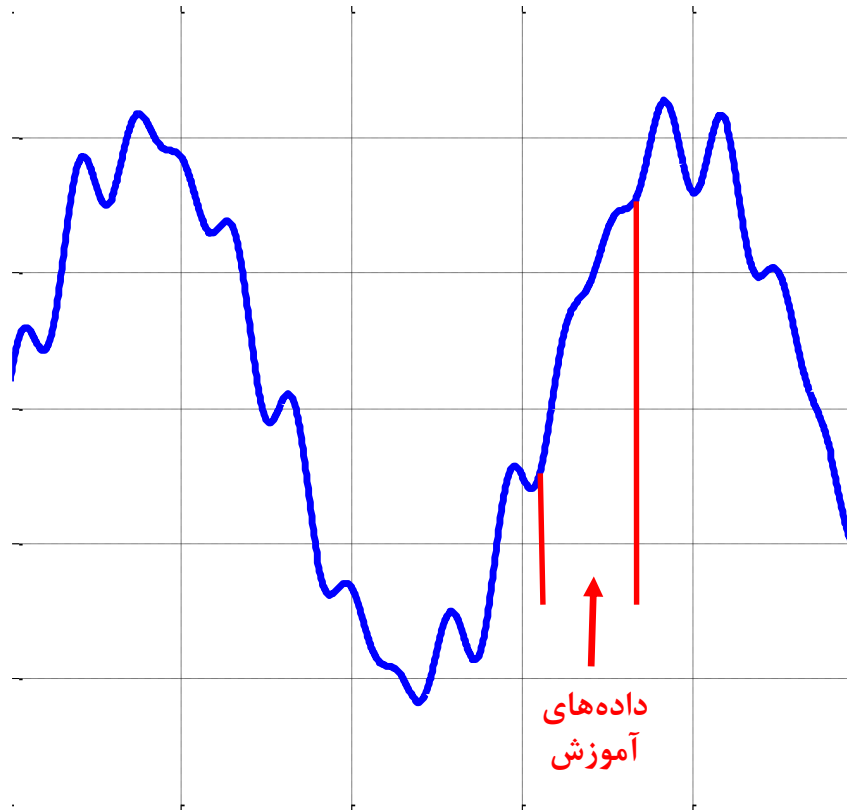
غنی بودن داده‌های آموزش:

– داده‌هایی که برای آموزش شبکه به کار می‌روند باید اطلاعات کافی از تابع غیرخطی موردنظر را در خود داشته باشد. ← غنی بودن داده‌ها

پرسپترون چندلایه (MLP)

غنی بودن داده‌های آموزش:

– داده‌هایی که برای آموزش شبکه به کار می‌روند باید اطلاعات کافی از تابع غیرخطی موردنظر را در خود داشته باشد. ← غنی بودن داده‌ها



مثال

پرسپترون چندلایه (MLP)

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

– یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.

پرسپترون چندلایه (MLP)

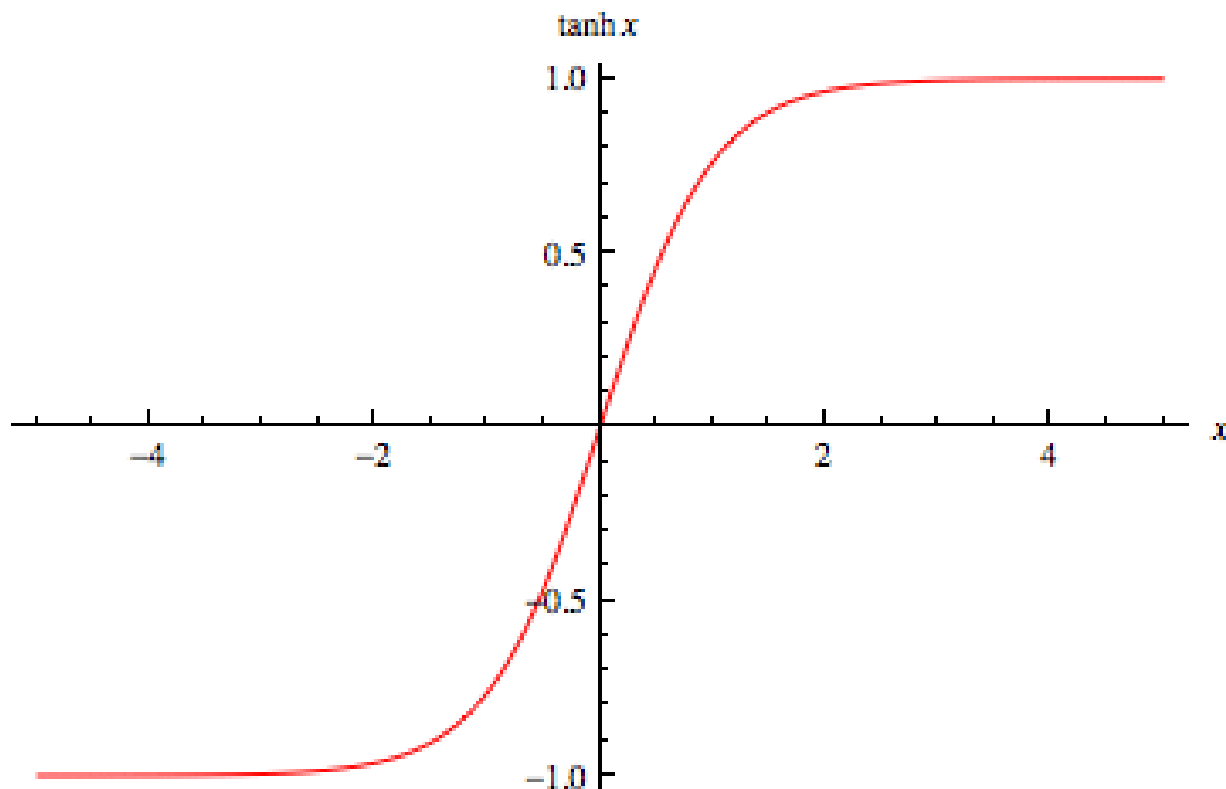
مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:
- ۱- اشباع خروجی سلول‌ها

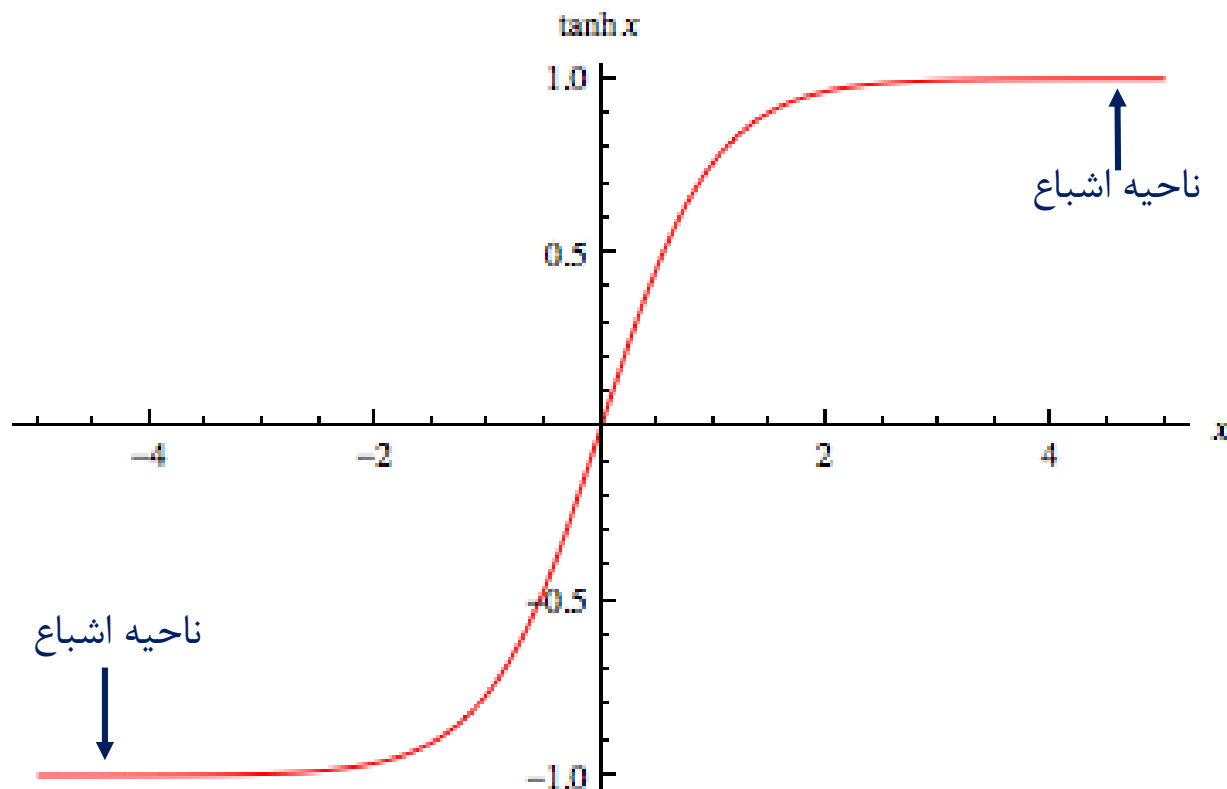


پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:

۱- اشباع خروجی سلول‌ها

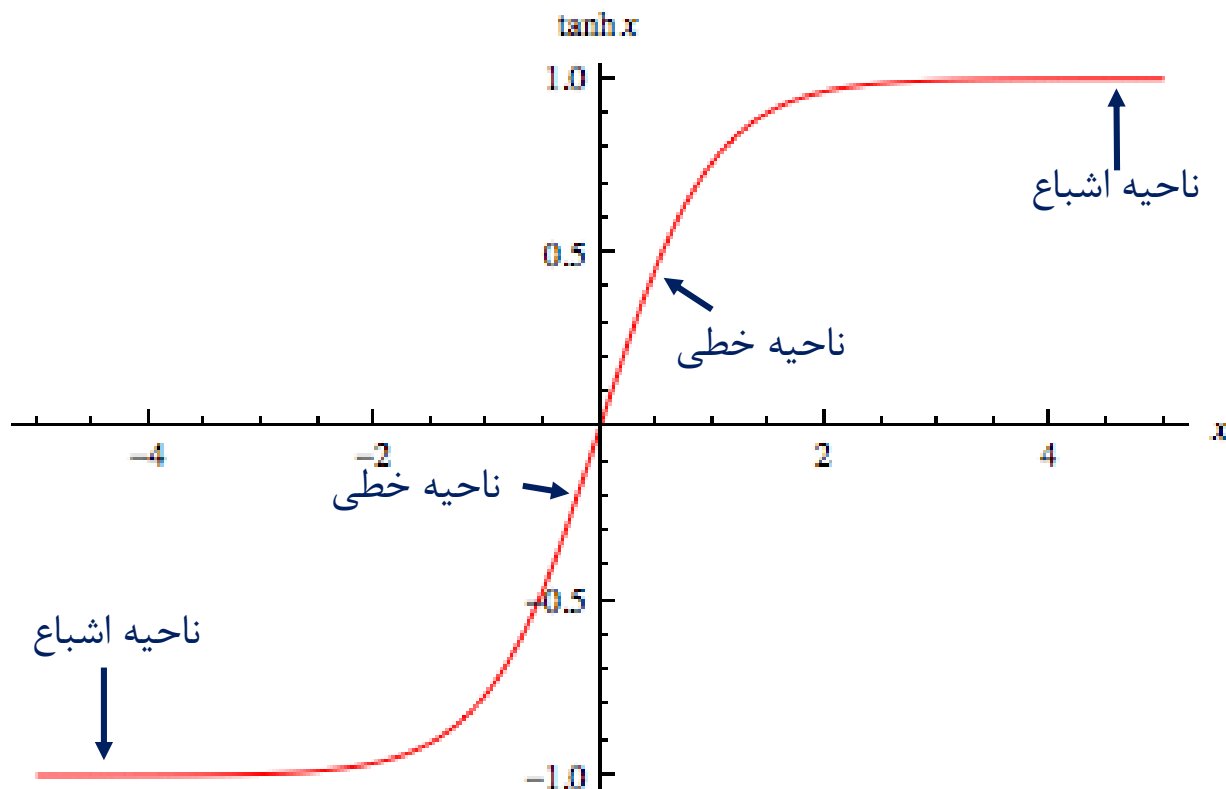


پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:

۱- اشباع خروجی سلول‌ها

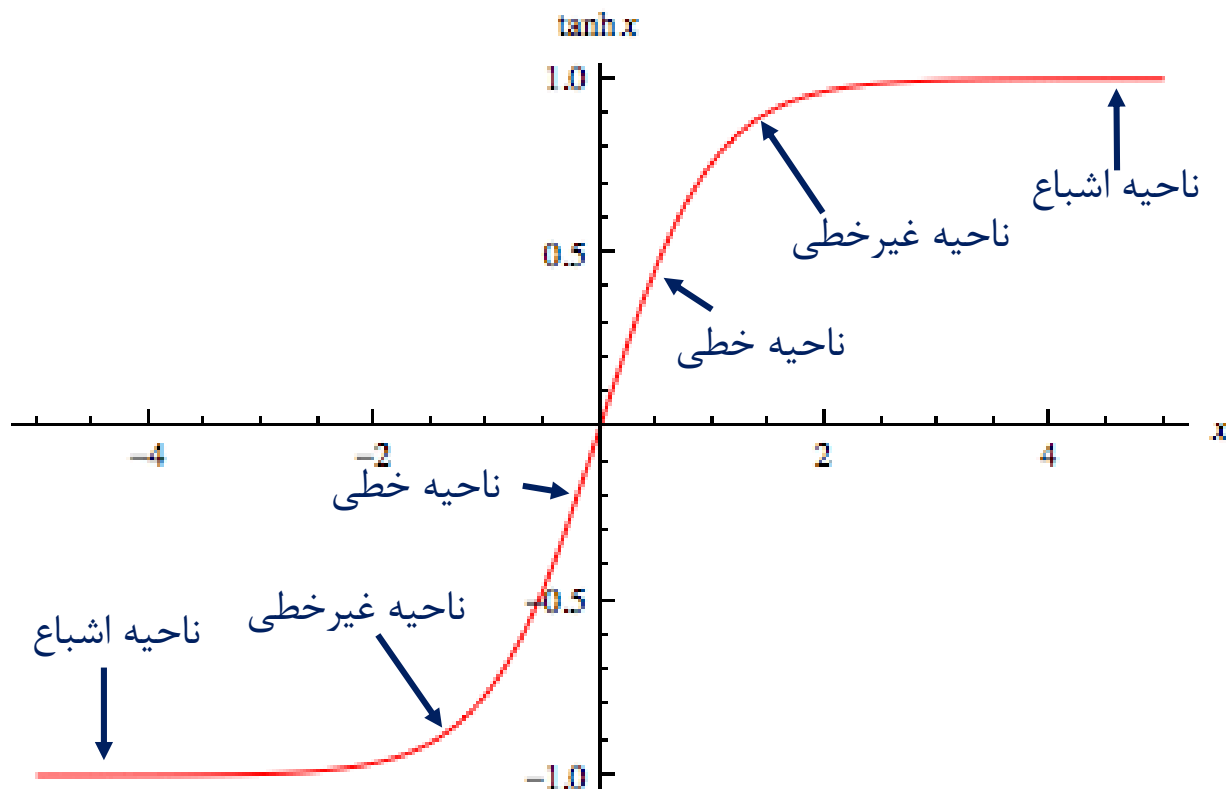


پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:

۱- اشباع خروجی سلول‌ها



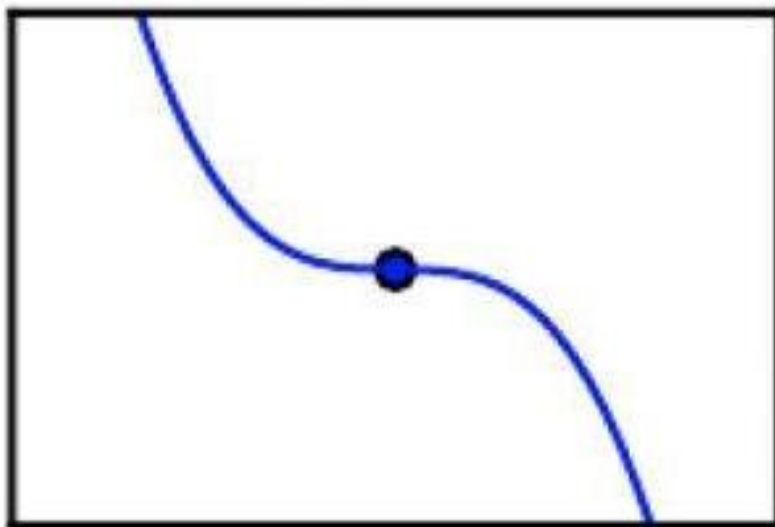
پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:

۱- اشباع خروجی سلول‌ها

۲- به‌وجود آمدن نقاط زینی (Saddle Points)



پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

- یکی از نکات مهم در الگوریتم پس‌انتشار خطا، انتخاب مقداری مناسب برای وزن‌های اولیه است.
- انتخاب نامناسب مقدار اولیه وزن‌ها می‌تواند باعث مشکلات زیر شود:
 - ۱- اشباع خروجی سلول‌ها
 - ۲- به‌وجود آمدن نقاط زینی (Saddle Points)
 - ۳- اشباع غلط

پرسپترون چندلایه (MLP)

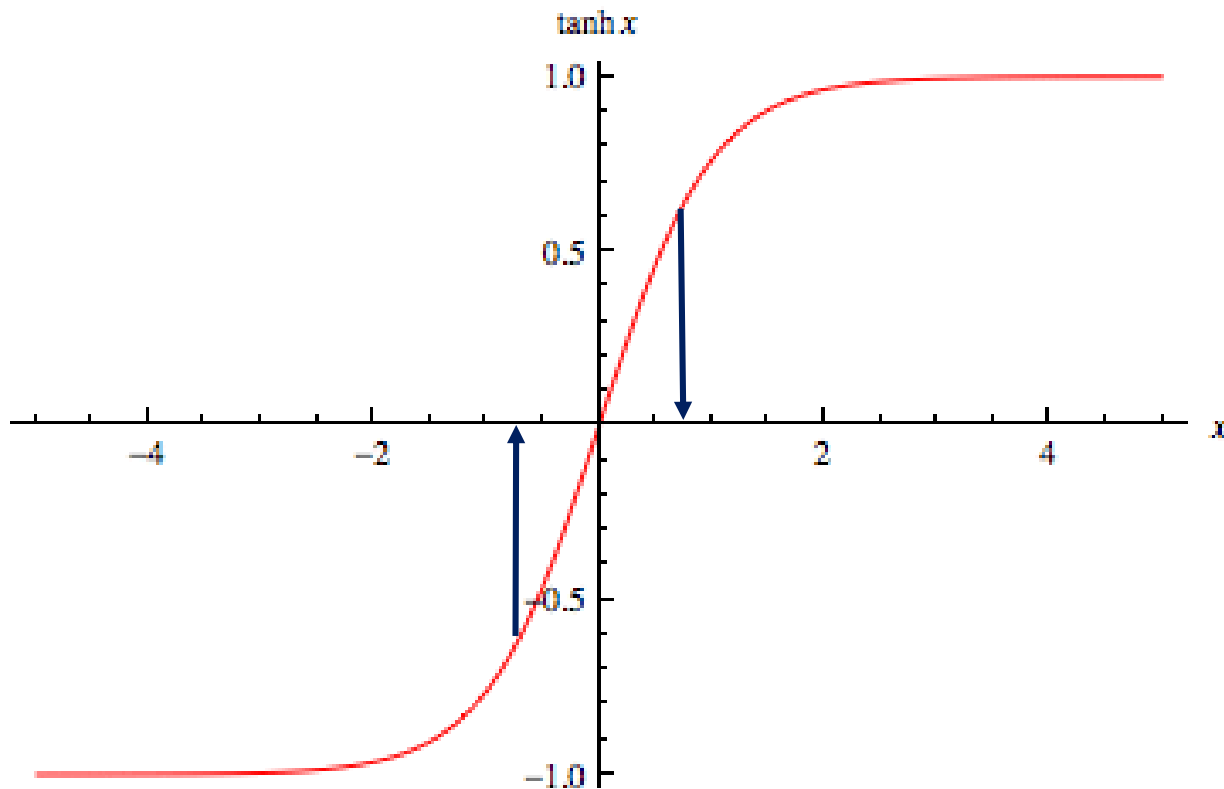
مقداردهی اولیه وزن‌ها:

– برای اجتناب از این مشکلات، باید مقدار اولیه وزن‌های طوری باشد که جمع خطی ورودی به سلول‌ها، در بین ناحیه‌های خطی و غیرخطی باشد.

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

– برای اجتناب از این مشکلات، باید مقدار اولیه وزن‌های طوری باشد که جمع خطی ورودی به سلول‌ها، در بین ناحیه‌های خطی و غیرخطی باشد.



پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

– برای به دست آوردن معادله برای این منظور

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

– برای به دست آوردن معادله برای این منظور

جمع خطی ورودی‌ها به سلول j

$$v_j = \sum_{i=1}^m w_{ji} y_i$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن ها:

– برای به دست آوردن معادله برای این منظور

$$v_j = \sum_{i=1}^m w_{ji} y_i \quad \text{جمع خطی ورودی ها به سلول } j$$

فرض کنید ورودی های هر سلول در شبکه دارای میانگین صفر و واریانس واحد باشد

$$\mu_y = \mathbb{E}[y_i] = 0 \quad \text{for all } i$$

$$\sigma_y^2 = \mathbb{E}[(y_i - \mu_i)^2] = \mathbb{E}[y_i^2] = 1 \quad \text{for all } i$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

– برای به دست آوردن معادله برای این منظور

$$v_j = \sum_{i=1}^m w_{ji} y_i \quad \text{جمع خطی ورودی‌ها به سلول } j$$

فرض کنید ورودی‌های هر سلول در شبکه دارای میانگین صفر و واریانس واحد باشد

$$\mu_y = \mathbb{E}[y_i] = 0 \quad \text{for all } i$$

$$\sigma_y^2 = \mathbb{E}[(y_i - \mu_i)^2] = \mathbb{E}[y_i^2] = 1 \quad \text{for all } i$$

همچنین، فرض کنید ورودی‌ها از نظر آماری مستقل از یکدیگر باشند

$$\mathbb{E}[y_i y_k] = \begin{cases} 1 & \text{for } k = i \\ 0 & \text{for } k \neq i \end{cases}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

علاوه بر آن، فرض کنید وزن‌ها با استفاده از توزیع یکنواخت اعداد با میانگین صفر ایجاد شده باشند

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

علاوه بر آن، فرض کنید وزن‌ها با استفاده از توزیع یکنواخت اعداد با میانگین صفر ایجاد شده باشند

$$\mu_w = \mathbb{E}[w_{ji}] = 0 \quad \text{for all } (j, i) \text{ pairs}$$

$$\sigma_w^2 = \mathbb{E}[(w_{ji} - \mu_w)^2] = \mathbb{E}[w_{ji}^2] \quad \text{for all } (j, i) \text{ pairs}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

علاوه بر آن، فرض کنید وزن‌ها با استفاده از توزیع یکنواخت اعداد با میانگین صفر ایجاد شده باشند

$$\mu_w = \mathbb{E}[w_{ji}] = 0 \quad \text{for all } (j, i) \text{ pairs}$$

$$\sigma_w^2 = \mathbb{E}[(w_{ji} - \mu_w)^2] = \mathbb{E}[w_{ji}^2] \quad \text{for all } (j, i) \text{ pairs}$$

بنابراین، میانگین v_j

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

علاوه بر آن، فرض کنید وزن‌ها با استفاده از توزیع یکنواخت اعداد با میانگین صفر ایجاد شده باشند

$$\mu_w = \mathbb{E}[w_{ji}] = 0 \quad \text{for all } (j, i) \text{ pairs}$$

$$\sigma_w^2 = \mathbb{E}[(w_{ji} - \mu_w)^2] = \mathbb{E}[w_{ji}^2] \quad \text{for all } (j, i) \text{ pairs}$$

بنابراین، میانگین v_j

$$\mu_v = \mathbb{E}[v_j] = \mathbb{E}\left[\sum_{i=1}^m w_{ji}y_i\right] = \sum_{i=1}^m \mathbb{E}[w_{ji}]\mathbb{E}[y_i] = 0$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

و واریانس v_j

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

و واریانس v_j

$$\sigma_v^2 = \mathbb{E}[(v_j - \mu_v)^2] = \mathbb{E}[v_j^2]$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن ها:

و واریانس v_j

$$\begin{aligned}\sigma_v^2 &= \mathbb{E}[(v_j - \mu_v)^2] = \mathbb{E}[v_j^2] \\ &= \mathbb{E}\left[\sum_{i=1}^m \sum_{k=1}^m w_{ji} w_{jk} y_i y_k\right]\end{aligned}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن ها:

و واریانس v_j

$$\begin{aligned}\sigma_v^2 &= \mathbb{E}[(v_j - \mu_v)^2] = \mathbb{E}[v_j^2] \\ &= \mathbb{E}\left[\sum_{i=1}^m \sum_{k=1}^m w_{ji} w_{jk} y_i y_k\right] \\ &= \sum_{i=1}^m \sum_{k=1}^m \mathbb{E}[w_{ji} w_{jk}] \mathbb{E}[y_i y_k]\end{aligned}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن ها:

و واریانس v_j

$$\begin{aligned}\sigma_v^2 &= \mathbb{E}[(v_j - \mu_v)^2] = \mathbb{E}[v_j^2] \\ &= \mathbb{E}\left[\sum_{i=1}^m \sum_{k=1}^m w_{ji} w_{jk} y_i y_k\right] \\ &= \sum_{i=1}^m \sum_{k=1}^m \mathbb{E}[w_{ji} w_{jk}] \mathbb{E}[y_i y_k] \\ &= \sum_{i=1}^m \mathbb{E}[w_{ji}^2]\end{aligned}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن ها:

و واریانس v_j

$$\begin{aligned}\sigma_v^2 &= \mathbb{E}[(v_j - \mu_v)^2] = \mathbb{E}[v_j^2] \\&= \mathbb{E}\left[\sum_{i=1}^m \sum_{k=1}^m w_{ji} w_{jk} y_i y_k\right] \\&= \sum_{i=1}^m \sum_{k=1}^m \mathbb{E}[w_{ji} w_{jk}] \mathbb{E}[y_i y_k] \\&= \sum_{i=1}^m \mathbb{E}[w_{ji}^2] \\&= m \sigma_w^2\end{aligned}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

در صورت استفاده از تابع تانژانت هیپربولیک به صورت زیر:

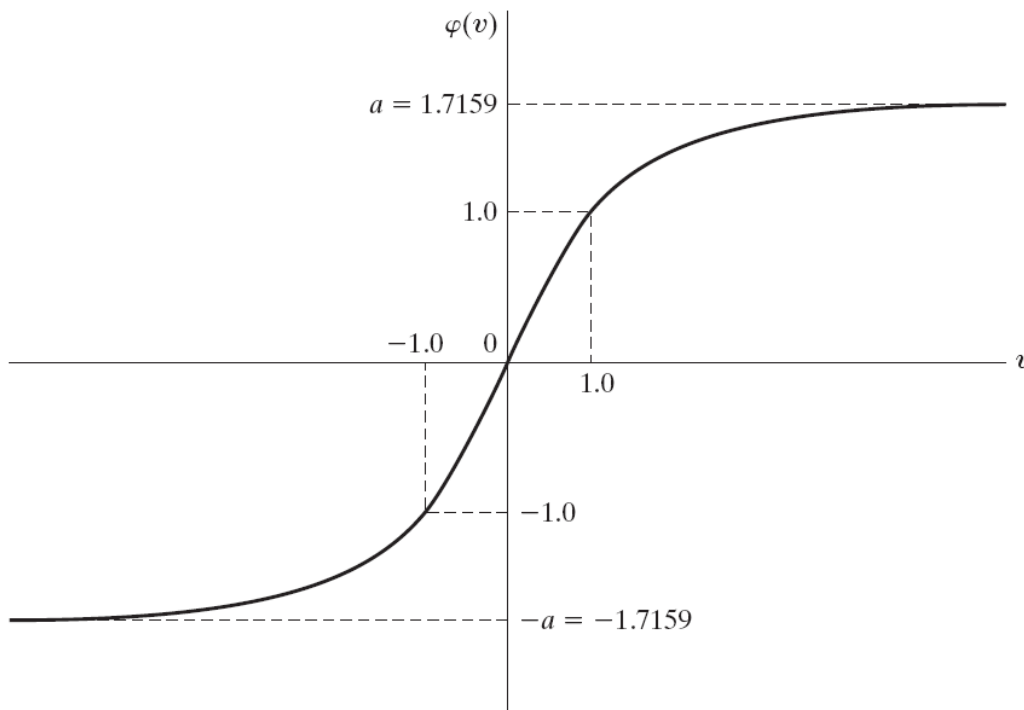
$$\varphi(v) = a \tanh(bv) \quad a = 1.7159 \quad b = \frac{2}{3}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

در صورت استفاده از تابع تانژانت هیپربولیک به صورت زیر:

$$\varphi(v) = a \tanh(bv) \quad a = 1.7159 \quad b = \frac{2}{3}$$

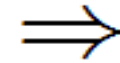
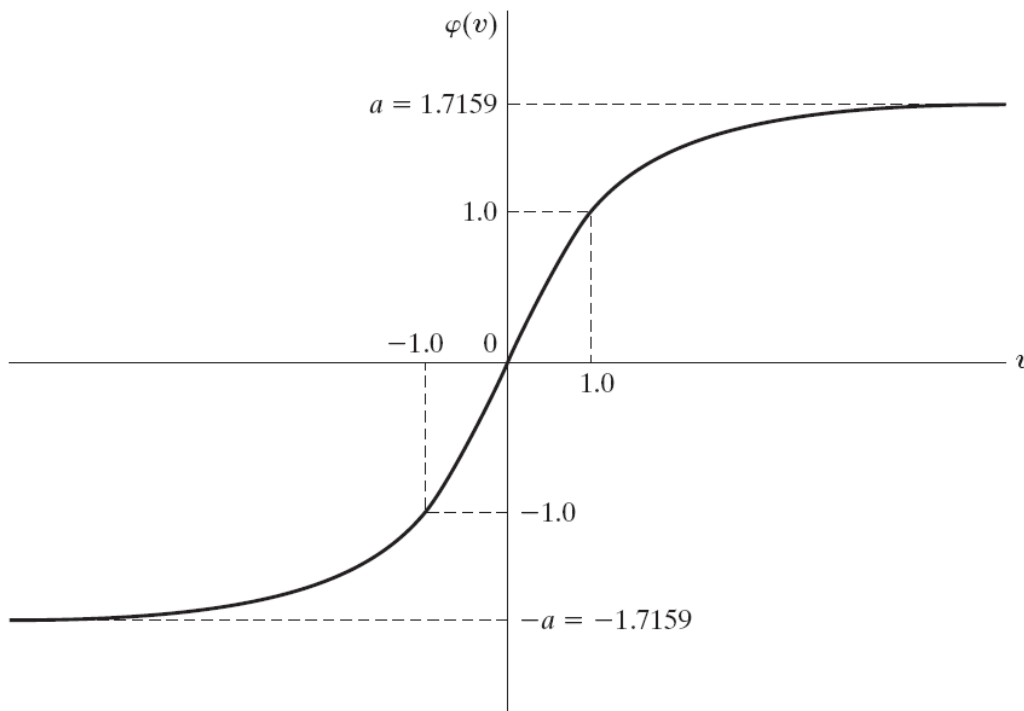


پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

در صورت استفاده از تابع تانژانت هیپربولیک به صورت زیر:

$$\varphi(v) = a \tanh(bv) \quad a = 1.7159 \quad b = \frac{2}{3}$$



$$\sigma_v = 1$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن ها:

$$\sigma_v^2 = m\sigma_w^2$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

$$\sigma_v^2 = m\sigma_w^2$$

در نتیجه

$$\sigma_w = m^{-1/2}$$

پرسپترون چندلایه (MLP)

مقداردهی اولیه وزن‌ها:

$$\sigma_v^2 = m\sigma_w^2$$

در نتیجه

$$\sigma_w = m^{-1/2}$$

یعنی مناسب است که مقدار اولیه وزن‌ها با توزیع یکنواخت، میانگین صفر و واریانس زیر تولید شوند:

$$\sigma_w = \frac{1}{\sqrt{m}}$$

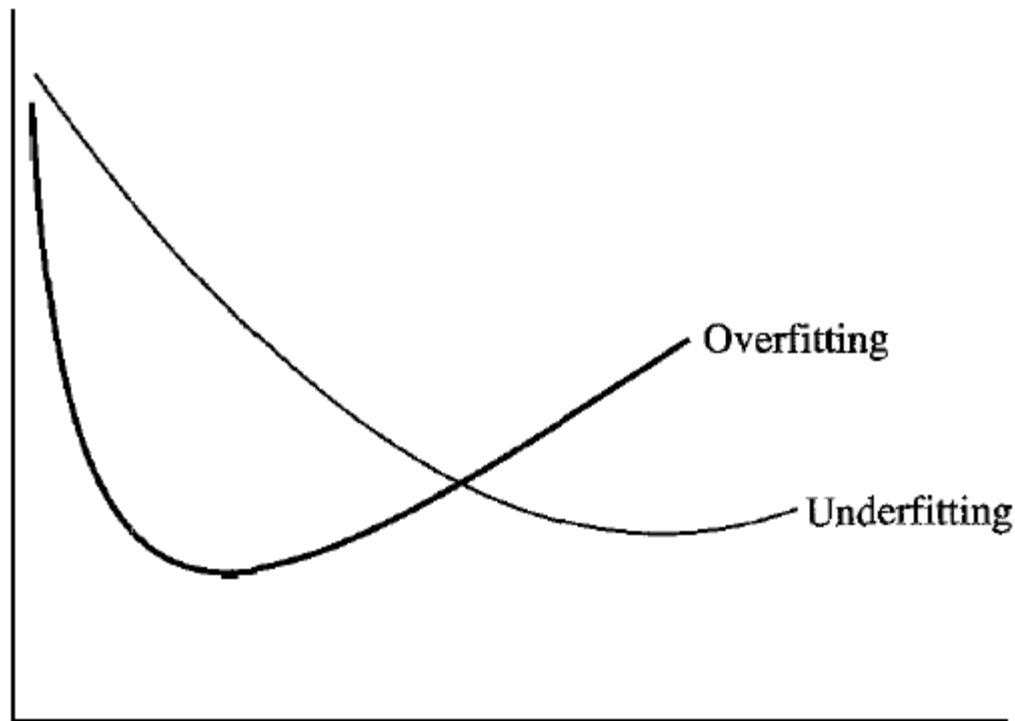
پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

برای جلوگیری از بیش‌برازش یا کم‌برازش در آموزش شبکه، باید تناسبی بین تعداد داده‌های آموزش و ساختار شبکه وجود داشته باشد.



پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

- ۱- تعداد نمونه‌های آموزش ثابت است و مساله تعیین ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) است. ← قبلا بررسی شد.

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

- ۱- تعداد نمونه‌های آموزش ثابت است و مساله تعیین ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) است. ← قبلا بررسی شد.
- ۲- ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) ثابت است و مساله تعیین تعداد نمونه‌های آموزش شبکه است. رابطه تقریبی برای این منظور

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

۱- تعداد نمونه‌های آموزش ثابت است و مساله تعیین ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) است. ← قبلا بررسی شد.

۲- ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) ثابت است و مساله تعیین تعداد نمونه‌های آموزش شبکه است. رابطه تقریبی برای این منظور

$$N \simeq \frac{W}{\varepsilon}$$

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

۱- تعداد نمونه‌های آموزش ثابت است و مساله تعیین ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) است. ← قبلا بررسی شد.

۲- ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) ثابت است و مساله تعیین تعداد نمونه‌های آموزش شبکه است. رابطه تقریبی برای این منظور

$$N \simeq \frac{W}{\varepsilon}$$

W تعداد وزن‌ها

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

۱- تعداد نمونه‌های آموزش ثابت است و مساله تعیین ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) است. ← قبلا بررسی شد.

۲- ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) ثابت است و مساله تعیین تعداد نمونه‌های آموزش شبکه است. رابطه تقریبی برای این منظور

$$N \simeq \frac{W}{\varepsilon}$$

W تعداد وزن‌ها

ε برابر بیشینه خطای مجاز به‌هنگام آزمایش شبکه ($0 < a < 1$)

پرسپترون چندلایه (MLP)

تعداد داده‌های لازم برای عمومیت دادن مناسب:

دو حالت را می‌توان در نظر گرفت:

۱- تعداد نمونه‌های آموزش ثابت است و مساله تعیین ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) است. ← قبلا بررسی شد.

۲- ساختار شبکه (تعداد لایه‌های پنهان و تعداد سلول در هر لایه) ثابت است و مساله تعیین تعداد نمونه‌های آموزش شبکه است. رابطه تقریبی برای این منظور

$$N \simeq \frac{W}{\varepsilon}$$

W تعداد وزن‌ها

ε برابر بیشینه خطای مجاز به‌هنگام آزمایش شبکه ($0 < a < 1$)

به‌عنوان مثال، برای $\varepsilon = 10\%$ ، تعداد نمونه‌های لازم تقریباً ده برابر تعداد وزن‌های شبکه است.

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.
- علاوه بر آن، مناسب است که هر وزنی، ضریب آموزش مختص خودش را داشته باشد. یعنی

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.
- علاوه بر آن، مناسب است که هر وزنی، ضریب آموزش مختص خودش را داشته باشد. یعنی

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta \eta_{kj}(n)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.
- علاوه بر آن، مناسب است که هر وزنی، ضریب آموزش مختص خودش را داشته باشد. یعنی

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta \eta_{kj}(n)$$

-
- نیاز به معادلات قبلی داریم

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.
- علاوه بر آن، مناسب است که هر وزنی، ضریب آموزش مختص خودش را داشته باشد. یعنی

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta \eta_{kj}(n)$$

-
- نیاز به معادلات قبلی داریم

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.
- علاوه بر آن، مناسب است که هر وزنی، ضریب آموزش مختص خودش را داشته باشد. یعنی

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta \eta_{kj}(n)$$

-
- نیاز به معادلات قبلی داریم

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n)$$

$$w_{kj}(n) = w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

پرسترون چندلایه (MLP)

تطبیق ضریب آموزش:

- یکی از راه‌های تسریع آموزش و در عین حال جلوگیری از واگرایی الگوریتم، تطبیق دادن ضریب آموزش است.
- علاوه بر آن، مناسب است که هر وزنی، ضریب آموزش مختص خودش را داشته باشد. یعنی

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta \eta_{kj}(n)$$

- نیاز به معادلات قبلی داریم

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n)$$

$$w_{kj}(n) = w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

$$\frac{\partial E(n)}{\partial w_{kj}(n)} = -e_k(n) \varphi'_k(v_k(n)) y_j(n)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- حال ضریب آموزش

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta\eta_{kj}(n)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- حال ضریب آموزش

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta\eta_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) - \gamma \frac{\partial E(n)}{\partial \eta_{kj}(n)}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- حال ضریب آموزش

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta\eta_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) - \gamma \frac{\partial E(n)}{\partial \eta_{kj}(n)}$$

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

تابع هزینه

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- حال ضریب آموزش

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta\eta_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) - \gamma \frac{\partial E(n)}{\partial \eta_{kj}(n)}$$

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

تابع هزینه

استفاده از قاعده مشتق زنجیره ای

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = \frac{\partial E(n)}{\partial y_k(n)} \frac{\partial y_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial \eta_{kj}(n)}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- حال ضریب آموزش

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta\eta_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) - \gamma \frac{\partial E(n)}{\partial \eta_{kj}(n)}$$

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

تابع هزینه

استفاده از قاعده مشتق زنجیره ای

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = \frac{\partial E(n)}{\partial y_k(n)} \frac{\partial y_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial \eta_{kj}(n)}$$
$$\downarrow \quad \quad \downarrow$$
$$-e_k(n) \quad \varphi'_k(v_k(n))$$

قبلا داشتیم

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

- حال ضریب آموزش

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \Delta\eta_{kj}(n)$$

$$\eta_{kj}(n+1) = \eta_{kj}(n) - \gamma \frac{\partial E(n)}{\partial \eta_{kj}(n)}$$

$$E(n) = \frac{1}{2} \sum_{k=1}^m e_k^2(n)$$

تابع هزینه

استفاده از قاعده مشتق زنجیره ای

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = \frac{\partial E(n)}{\partial y_k(n)} \frac{\partial y_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial \eta_{kj}(n)}$$
$$\downarrow \quad \quad \downarrow \quad \quad \downarrow$$
$$-e_k(n) \quad \varphi'_k(v_k(n)) \quad ?$$

قبلا داشتیم

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial v_k(n)}{\partial \eta_{kj}(n)} \text{ محاسبه}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial v_k(n)}{\partial \eta_{kj}(n)} \text{ محاسبه}$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (1)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial v_k(n)}{\partial \eta_{kj}(n)} \text{ محاسبه}$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (۱)$$

$$w_{kj}(n) = w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \quad (۲)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial v_k(n)}{\partial \eta_{kj}(n)} \text{ محاسبه}$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (1)$$

$$w_{kj}(n) = w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \quad (2)$$

(2) در (1)

$$v_k(n) = \sum_{j=0}^q y_j(n) \left[w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \right]$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial v_k(n)}{\partial \eta_{kj}(n)} \text{ محاسبه}$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (1)$$

$$w_{kj}(n) = w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \quad (2)$$

(2) در (1)

$$v_k(n) = \sum_{j=0}^q y_j(n) \left[w_{kj}(n-1) - \eta_{kj}(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \right]$$

در نتیجه

$$\frac{\partial v_k(n)}{\partial \eta_{kj}(n)} = -y_j(n) \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = -e_k(n) \varphi'_k(v_k(n)) y_j(n) \left[\frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \right] \quad \text{بنابراین}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = \underbrace{-e_k(n) \varphi'_k(v_k(n)) y_j(n)}_{\frac{\partial E(n)}{\partial w_{kj}(n)}} \left[\frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \right] \quad \text{بنابراین}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = \underbrace{-e_k(n) \varphi'_k(v_k(n)) y_j(n)}_{\frac{\partial E(n)}{\partial w_{kj}(n)}} \left[\frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \right] \quad \text{بنابراین}$$

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = - \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \quad \text{در نتیجه}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = \underbrace{-e_k(n) \varphi'_k(v_k(n)) y_j(n)}_{\frac{\partial E(n)}{\partial w_{kj}(n)}} \left[\frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \right] \quad \text{بنابراین}$$

$$\frac{\partial E(n)}{\partial \eta_{kj}(n)} = - \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \quad \text{در نتیجه}$$

$$\begin{aligned} \Delta \eta_{kj}(n) &= -\gamma \frac{\partial E(n)}{\partial \eta_{kj}(n)} \\ &= \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)} \quad \text{در نهایت} \end{aligned}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه گیری

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه‌گیری

۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می‌یابد. ← تسریع در همگرایی

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه‌گیری

- ۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می‌یابد. ← تسریع در همگرایی
- ۲- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت متفاوت باشد، ضریب آموزش کاهش می‌یابد. ← پایداری الگوریتم

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه‌گیری

- ۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می‌یابد. ← تسریع در همگرایی
- ۲- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت متفاوت باشد، ضریب آموزش کاهش می‌یابد. ← پایداری الگوریتم

به این روش، الگوریتم دلتا-دلتا می‌گویند

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه گیری

- ۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می یابد. ← تسریع در همگرایی
- ۲- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت متفاوت باشد، ضریب آموزش کاهش می یابد. ← پایداری الگوریتم

به این روش، الگوریتم دلتا-دلتا می گویند

معایب این روش:

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه گیری

- ۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می یابد. ← تسریع در همگرایی
- ۲- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت متفاوت باشد، ضریب آموزش کاهش می یابد. ← پایداری الگوریتم

به این روش، الگوریتم دلتا-دلتا می گویند

معایب این روش:

- نیاز به حافظه بیشتر نسبت به الگوریتم پس انتشار خطا

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه گیری

- ۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می یابد. ← تسریع در همگرایی
- ۲- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت متفاوت باشد، ضریب آموزش کاهش می یابد. ← پایداری الگوریتم

به این روش، الگوریتم دلتا-دلتا می گویند

معایب این روش:

- نیاز به حافظه بیشتر نسبت به الگوریتم پس انتشار خطا
- یافتن مقدار مناسبی برای گاما

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\eta_{kj}(n+1) = \eta_{kj}(n) + \gamma \frac{\partial E(n)}{\partial w_{kj}(n)} \frac{\partial E(n-1)}{\partial w_{kj}(n-1)}$$

نتیجه گیری

- ۱- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت یکسان باشد، ضریب آموزش افزایش می یابد. ← تسریع در همگرایی
- ۲- هنگامی که مشتق تابع هزینه نسبت به وزن مورد نظر در دو مرحله پیاپی دارای علامت متفاوت باشد، ضریب آموزش کاهش می یابد. ← پایداری الگوریتم

به این روش، الگوریتم دلتا-دلتا می گویند

معایب این روش:

- نیاز به حافظه بیشتر نسبت به الگوریتم پس انتشار خطا
- یافتن مقدار مناسبی برای گاما
- امکان منفی شدن ضریب آموزش

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

برای رفع عیب سوم الگوریتم دلتا-دلتا، از علامت و مقدار مشتق در تمامی مراحل قبلی استفاده می‌کنیم

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

برای رفع عیب سوم الگوریتم دلتا-دلتا، از علامت و مقدار مشتق در تمامی مراحل قبلی استفاده می‌کنیم

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

برای رفع عیب سوم الگوریتم دلتا-دلتا، از علامت و مقدار مشتق در تمامی مراحل قبلی استفاده می‌کنیم

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

که در آن

$$D_{kj}(n) = \frac{\partial E(n)}{\partial w_{kj}(n)}$$

$$S_{kj}(n) = (1 - \xi) D_{kj}(n-1) + \xi S_{kj}(n-1)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

برای رفع عیب سوم الگوریتم دلتا-دلتا، از علامت و مقدار مشتق در تمامی مراحل قبلی استفاده می‌کنیم

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

که در آن

$$D_{kj}(n) = \frac{\partial E(n)}{\partial w_{kj}(n)}$$

$$S_{kj}(n) = (1 - \xi) D_{kj}(n-1) + \xi S_{kj}(n-1)$$

با حل معادله گسسته اخیر

$$S_{kj}(n) = \sum_{p=1}^{n-2} \xi^p (1 - \xi) D_{kj}(n-p-1) + \xi^{(n-1)} S_{kj}(n-1)$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

برای رفع عیب سوم الگوریتم دلتا-دلتا، از علامت و مقدار مشتق در تمامی مراحل قبلی استفاده می‌کنیم

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

که در آن

$$D_{kj}(n) = \frac{\partial E(n)}{\partial w_{kj}(n)}$$

$$S_{kj}(n) = (1 - \xi) D_{kj}(n-1) + \xi S_{kj}(n-1)$$

با حل معادله گسسته اخیر

$$S_{kj}(n) = \sum_{p=1}^{n-2} \xi^p (1 - \xi) D_{kj}(n-p-1) + \xi^{(n-1)} S_{kj}(n-1)$$

در این سری زمانی، مشتق تابع هزینه در تمامی مراحل قبلی برای تغییر ضریب آموزش در نظر گرفته می‌شود.

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

۱- با قراردادن لاندا و بتا برابر صفر ← ؟

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

۱- با قراردادن لاندا و بتا برابر صفر ← الگوریتم پس انتشار خطا

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

۱- با قراردادن لاندا و بتا برابر صفر ← الگوریتم پس انتشار خطا

۲- افزایش در ضریب آموزش (یعنی حالتی که مشتق در تمامی مراحل قبلی دارای علامت یکسان باشد) به طور خطی است.

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

- ۱- با قراردادن لاندا و بتا برابر صفر ← الگوریتم پس انتشار خطا
- ۲- افزایش در ضریب آموزش (یعنی حالتی که مشتق در تمامی مراحل قبلی دارای علامت یکسان باشد) به طور خطی است.
- ۳- کاهش در ضریب آموزش (یعنی حالتی که مشتق در مراحل قبلی دارای علامت متفاوت باشد) به طور نمایی است.

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

- ۱- با قراردادن لاندا و بتا برابر صفر ← الگوریتم پس انتشار خطا
- ۲- افزایش در ضریب آموزش (یعنی حالتی که مشتق در تمامی مراحل قبلی دارای علامت یکسان باشد) به طور خطی است.
- ۳- کاهش در ضریب آموزش (یعنی حالتی که مشتق در مراحل قبلی دارای علامت متفاوت باشد) به طور نمایی است.
- ۴- ضریب آموزش همواره مثبت باقی می ماند.

پرسپترون چندلایه (MLP)

تطبیق ضریب آموزش:

$$\Delta \eta_{kj}(n) = \begin{cases} \lambda & \text{if } S_{kj}(n-1)D_{kj}(n) > 0 \\ -\beta \eta_{kj}(n) & \text{if } S_{kj}(n-1)D_{kj}(n) < 0 \\ 0 & \text{otherwise} \end{cases}$$

در این روش

- ۱- با قراردادن لاندا و بتا برابر صفر ← الگوریتم پس انتشار خطا
- ۲- افزایش در ضریب آموزش (یعنی حالتی که مشتق در تمامی مراحل قبلی دارای علامت یکسان باشد) به طور خطی است.
- ۳- کاهش در ضریب آموزش (یعنی حالتی که مشتق در مراحل قبلی دارای علامت متفاوت باشد) به طور نمایی است.
- ۴- ضریب آموزش همواره مثبت باقی می ماند.

به این روش، الگوریتم دلتا-بار-دلتا می گویند.

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

در منطق فازی:

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

در منطق فازی:

مثال:

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریبا برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریباً برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

همچنین، فرض کنید که گرادیان دوم سطح خط تقریباً برابر با تغییرات در تغییرات خطا باشد و آن را با CCE نشان دهیم

$$CCE(n) = CE(n) - CE(n - 1)$$

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریباً برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

همچنین، فرض کنید که گرادیان دوم سطح خطا تقریباً برابر با تغییرات در تغییرات خطا باشد و آن را با CCE نشان دهیم

$$CCE(n) = CE(n) - CE(n - 1)$$

در این صورت، موارد زیر را می‌توان تشخیص داد:

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریباً برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

همچنین، فرض کنید که گرادیان دوم سطح خط تقریباً برابر با تغییرات در تغییرات خطا باشد و آن را با CCE نشان دهیم

$$CCE(n) = CE(n) - CE(n - 1)$$

در این صورت، موارد زیر را می‌توان تشخیص داد:

۱- اگر CE کوچک باشد و هیچ تغییر علامتی در چندین مرحله متوالی تکرار وجود نداشته باشد، آنگاه ضریب یادگیری را باید ...؟

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریباً برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

همچنین، فرض کنید که گرادیان دوم سطح خطا تقریباً برابر با تغییرات در تغییرات خطا باشد و آن را با CCE نشان دهیم

$$CCE(n) = CE(n) - CE(n - 1)$$

در این صورت، موارد زیر را می‌توان تشخیص داد:

۱- اگر CE کوچک باشد و هیچ تغییر علامتی در چندین مرحله متوالی تکرار وجود نداشته باشد، آنگاه ضریب یادگیری را باید افزایش داد.

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریباً برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

همچنین، فرض کنید که گرادیان دوم سطح خط تقریباً برابر با تغییرات در تغییرات خطا باشد و آن را با CCE نشان دهیم

$$CCE(n) = CE(n) - CE(n - 1)$$

در این صورت، موارد زیر را می‌توان تشخیص داد:

۱- اگر CE کوچک باشد و هیچ تغییر علامتی در چندین مرحله متوالی تکرار وجود نداشته باشد، آنگاه ضریب یادگیری را باید افزایش داد.

۲- اگر در چندین مرحله تکرار متوالی، تغییر علامت در CE وجود داشته باشد، در آنگاه ضریب یادگیری را باید ... ؟

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

در منطق فازی: IF <a set of conditions are satisfied>,
THEN <a set of consequences are inferred>

مثال:

فرض کنید که گرادیان سطح خط تقریبا برابر با تغییرات در خطا (منظور جمع مربعات لحظه‌ای خطا) باشد و آن را با CE نشان دهیم

$$CE(n) = E(n) - E(n - 1)$$

همچنین، فرض کنید که گرادیان دوم سطح خطا تقریبا برابر با تغییرات در تغییرات خطا باشد و آن را با CCE نشان دهیم

$$CCE(n) = CE(n) - CE(n - 1)$$

در این صورت، موارد زیر را می‌توان تشخیص داد:

۱- اگر CE کوچک باشد و هیچ تغییر علامتی در چندین مرحله متوالی تکرار وجود نداشته باشد، آنگاه ضریب یادگیری را باید افزایش داد.

۲- اگر در چندین مرحله تکرار متوالی، تغییر علامت در CE وجود داشته باشد، در آنگاه ضریب یادگیری را باید کاهش داد.

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

مثال (ادامه):

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

مثال (ادامه):

۳- اگر CE و CCE هر دو کوچک باشند و هیچ تغییر علامتی در چندین مرحله تکرار متوالی وجود نداشته باشد، در این صورت هر دو ضریب یادگیری و ممنتم را باید افزایش داد.

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

مثال (ادامه):

۳- اگر CE و CCE هر دو کوچک باشند و هیچ تغییر علامتی در چندین مرحله تکرار متوالی وجود نداشته باشد، در این صورت هر دو ضریب یادگیری و ممنتم را باید افزایش داد.

نتیجه گیری:

- به سه مورد بالا، قواعد فازی برای تغییر ضریب آموزش (و/یا ضریب ممنتم) می گویند.

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

مثال (ادامه):

۳- اگر CE و CCE هر دو کوچک باشند و هیچ تغییر علامتی در چندین مرحله تکرار متوالی وجود نداشته باشد، در این صورت هر دو ضریب یادگیری و ممنتم را باید افزایش داد.

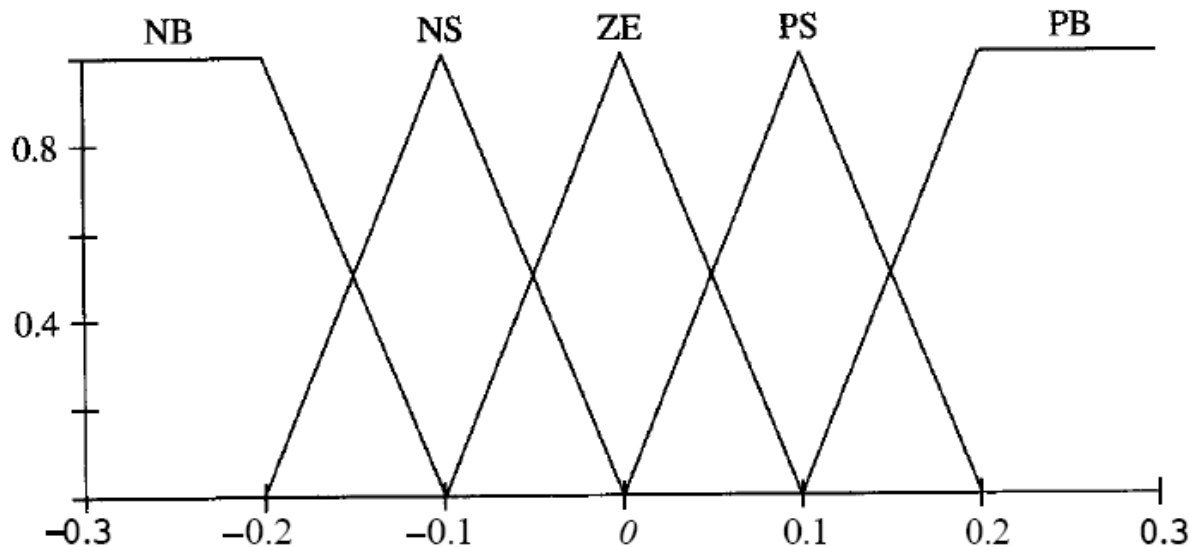
نتیجه گیری:

- به سه مورد بالا، قواعد فازی برای تغییر ضریب آموزش (و/یا ضریب ممنتم) می گویند.

سوال: فازی (مشکک) بودن این قواعد در چیست؟

پرسپترون چندلایه (MLP)

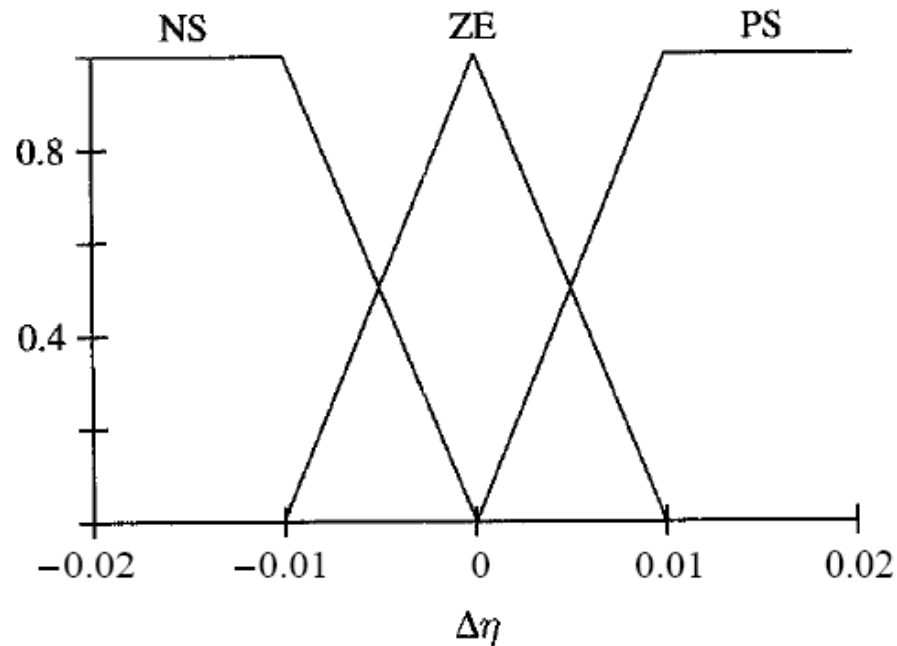
استفاده از منطق فازی در تطبیق ضریب آموزش:



توابع عضویت برای ورودی‌ها (CE و CCE)

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:



توابع عضویت برای خروجی $\Delta\eta$

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

- جدول قواعد نمونه برای تنظیم ضریب آموزش می تواند به صورت زیر باشد:

جدول قواعد فازی برای تعیین $\Delta\eta$

CCE	CE				
	NB	NS	ZE	PS	PB
NB	NS	NS	NS	NS	NS
NS	NS	ZE	PS	ZE	NS
ZE	ZE	PS	PS	PS	ZE
PS	NS	ZE	PS	ZE	NS
PB	NS	NS	NS	NS	NS

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

– جدول قواعد مشابهی را می توان برای تغییر ضریب ممنتیم نیز تعریف کرد.

جدول قواعد فازی برای تعیین $\Delta\alpha$

CCE	CE				
	NB	NS	ZE	PS	PB
NB	NS	NS	ZE	ZE	ZE
NS	NS	ZE	ZE	ZE	ZE
ZE	ZE	PS	PS	PS	ZE
PS	ZE	ZE	ZE	ZE	NS
PB	ZE	ZE	ZE	NS	NS

پرسپترون چندلایه (MLP)

استفاده از منطق فازی در تطبیق ضریب آموزش:

– در زمینه استفاده از شبکه‌های عصبی و منطق فازی مقالات و کتب فراوانی به چاپ رسیده است.

