Batch Norm

CLASS `torch.nn.BatchNorm2d(`*`num_features`*, *`eps=1e-05`*, *`momentum=0.1`*, *`affine=True`*, *`track_running_stats=True`*, *`device=None`*, *`dtype=None`*`)` [SOURCE]

Applies Batch Normalization over a 4D input (a mini-batch of 2D inputs with additional channel dimension) as described in the paper Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift .

$$y = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} * \gamma + \beta$$

The mean and standard-deviation are calculated per-dimension over the mini-batches and γ and β are learnable parameter vectors of size *C* (where *C* is the input size). By default, the elements of γ are set to 1 and the elements of β are set to 0. The standard-deviation is calculated via the biased estimator, equivalent to *torch.var*(*input, unbiased=False*).

Also by default, during training this layer keeps running estimates of its computed mean and variance, which are then used for normalization during evaluation. The running estimates are kept with a default `momentum` of 0.1.

If `track_running_stats` is set to `False`, this layer then does not keep running estimates, and batch statistics are instead used during evaluation time as well.

• NOTE

This `momentum` argument is different from one used in optimizer classes and the conventional notion of momentum. Mathematically, the update rule for running statistics here is $\hat{x}_{\text{new}} = (1 - \text{momentum}) \times \hat{x} + \text{momentum} \times x_t$, where $\hat{x}$ is the estimated statistic and x_t is the new observed value.

Because the Batch Normalization is done over the *C* dimension, computing statistics on (*N, H, W*) slices, it's common terminology to call this Spatial Batch Normalization.

SVD

Singular value decomposition

From Wikipedia, the free encyclopedia

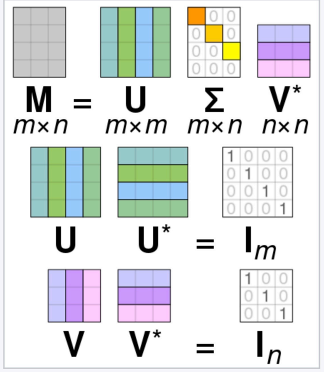
In linear algebra, the **singular value decomposition** (**SVD**) is a factorization of a real or complex matrix. It generalizes the eigendecomposition of a square normal matrix with an orthonormal eigenbasis to any $m \times n$ matrix. It is related to the polar decomposition.

Specifically, the singular value decomposition of an $m \times n$ complex matrix $\mathbf{M}$ is a factorization of the form $\mathbf{M} = \mathbf{U\Sigma V^*}$, where $\mathbf{U}$ is an $m \times m$ complex unitary matrix, $\mathbf{\Sigma}$ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and $\mathbf{V}$ is an $n \times n$ complex unitary matrix. If $\mathbf{M}$ is real, $\mathbf{U}$ and $\mathbf{V}$ can also be guaranteed to be real orthogonal matrices. In such contexts, the SVD is often denoted $\mathbf{U\Sigma V^T}$.

The diagonal entries $\sigma_i = \Sigma_{ii}$ of $\mathbf{\Sigma}$ are uniquely determined by $\mathbf{M}$ and are known as the singular values of $\mathbf{M}$. The number of non-zero singular values is equal to the rank of $\mathbf{M}$. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called left-singular vectors and right-singular vectors of $\mathbf{M}$, respectively. They form two sets of orthonormal bases $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and the singular value decomposition can be written as $\mathbf{M} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$, where $r \leq \min\{m, n\}$ is the rank of $\mathbf{M}$.

The SVD is not unique. It is always possible to choose the decomposition so that the singular values Σ_{ii} are in descending order. In this case, $\mathbf{\Sigma}$ (but not $\mathbf{U}$ and $\mathbf{V}$) is uniquely determined by $\mathbf{M}$.

The term sometimes refers to the **compact SVD**, a similar decomposition $\mathbf{M} = \mathbf{U\Sigma V^*}$ in which $\mathbf{\Sigma}$ is square diagonal of size $r \times r$, where $r \leq \min\{m, n\}$ is the rank of $\mathbf{M}$, and has only the non-zero singular values. In this variant, $\mathbf{U}$ is an $m \times r$ semi-unitary matrix and $\mathbf{V}$ is an $n \times r$ semi-unitary matrix, such that $\mathbf{U^*U} = \mathbf{V^*V} = \mathbf{I}_r$.



Visualization of the matrix multiplications in singular value decomposition