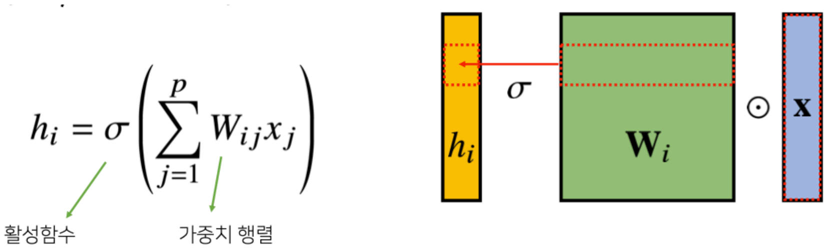
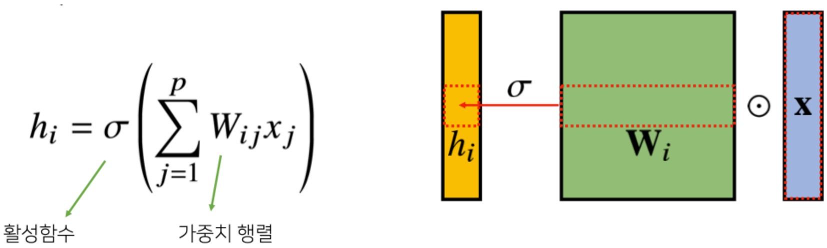
CNN

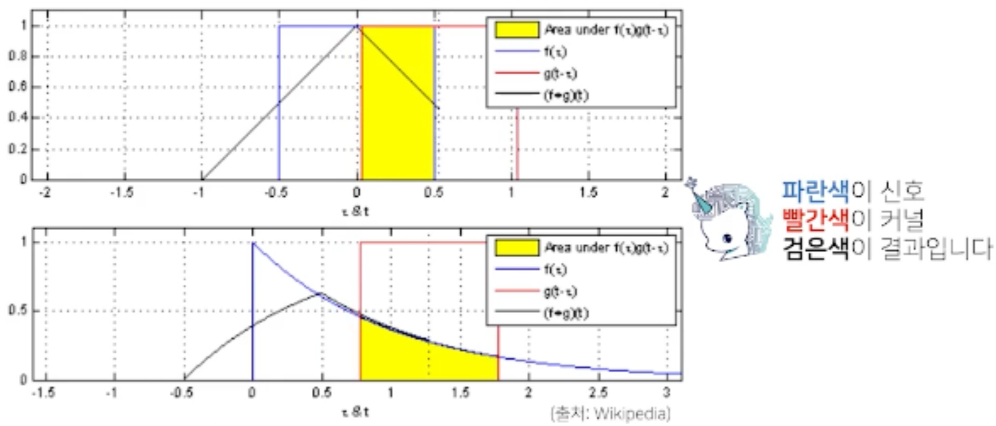
- **다층신경망(MLP)**: 각 뉴런들이 선형모델과 활성함수로 모두 연결된 (fully connected) 구조



- **Convolution 연산**: 커널(kernel)을 입력벡터 상에서 움직여가면서 선형모델과 합성함수가 적용되는 구조



- 활성화 함수를 제외한 convolution 연산도 선형변환에 속함
- 공통 커널 사용, 커널 사이즈 고정, 파라미터 사이즈 많이 줄일 수 있음
- Convolution 연산의 수학적인 의미: signal를 **커널을 이해 국소적으로 증폭/감소**시켜서 **정보를 추출 또는 필터링**하는 것
  - continuous: $[f*g](x) = \int_{\mathbb{R}^d} f(z)g(x-z)\mathrm{d}z = \int_{\mathbb{R}^d} f(x-z)g(z)\mathrm{d}z = [g*f](x)$
  - discrete: $[f*g](i) = \sum_{a\in\mathbb{Z}^d} f(a)g(i-a) = \sum_{a\in\mathbb{Z}^d} f(i-a)g(a) = [g*f](i)$
- CNN에서 사용하는 연산은 사실 **convolution**이 아니라 **cross-correlation**이라고 부른다.
  - continuous: $[f*g](x) = \int_{\mathbb{R}^d} f(z)g(x+z)\mathrm{d}z = \int_{\mathbb{R}^d} f(x+z)g(z)\mathrm{d}z = [g*f](x)$
  - discrete: $[f*g](i) = \sum_{a\in\mathbb{Z}^d} f(a)g(i+a) = \sum_{a\in\mathbb{Z}^d} f(i+a)g(a) = [g*f](i)$
- 커널은 정의역 내에서 움직여도 변화하지 않고 **(translation invariant)** 주어진 신호에 **국소적(local)**으로 적용

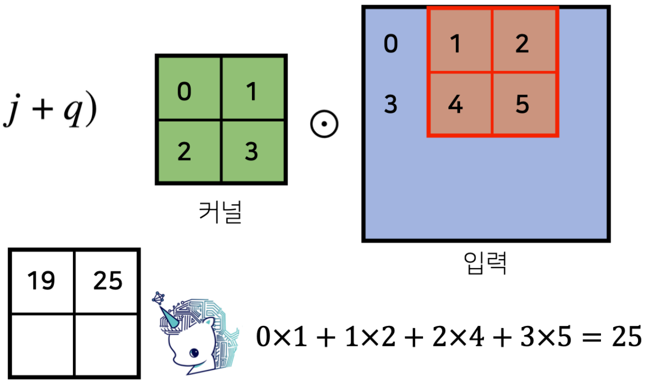


- 다양한 차원에서 계산 가능
  - 1D-conv: $[f*g](i) = \sum_{p=1}^{d} f(p)g(i+p)$
  - 2D-conv: $[f*g](i,j) = \sum_{p,q} f(p,q)g(i+p,j+q)$
  - 3D-conv: $[f*g](i,j,k) = \sum_{p,q,r} f(p,q,r)g(i+p,j+q,k+r)$

  i, j, k가 바뀌어도 커널 f의 값은 바뀌지 않습니다.

- **2차원 Convolution 연산 이해하기**
  - 2D-Conv 연산은 이와 달리 커널(kernel)을 입력벡터 상에서 움직여가면서 선형모델과 합성함수가 적용되는 구조

$$[f * g](i, j) = \sum_{p,q} \overbrace{f(p,q)}^{\text{커널}} \underbrace{g(i+p, j+q)}_{\text{입력}}$$



  - 입력 크기를 (H, W), 커널 크기를 (K_H, K_W), 출력 크기를 (O_H, O_W)라 하면 출력 크기는 다음과 같이 계산함

    $O_H = H - K_H + 1$

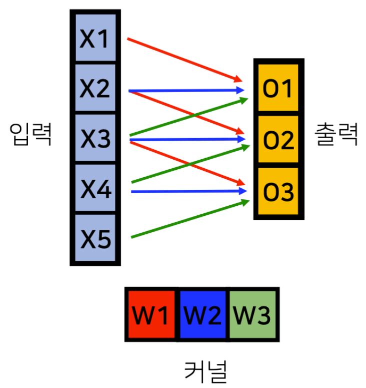
    $O_W = W - K_W + 1$
    - ex) 28X28 입력을 3X3 커널로 2D-Conv 연산을 하면 26X26이 된다.
- 채널이 여러개 2차원 입력의 경우, 2차원 Convolution을 채널 개수만큼 적용한다고 생각하기
- 3차원부터는 행렬이 아닌 **텐서**라고 부름
- 채널이 여러개인 경우, 커널의 채널 수와 입력의 채널 수가 같아야 함
- 텐서 => 직육면체 블록으로 이해하면 좀 더 이해하기 쉬움

  $(K_H, K_W, C) * (H, W, C) \to (O_H, O_W, 1)$

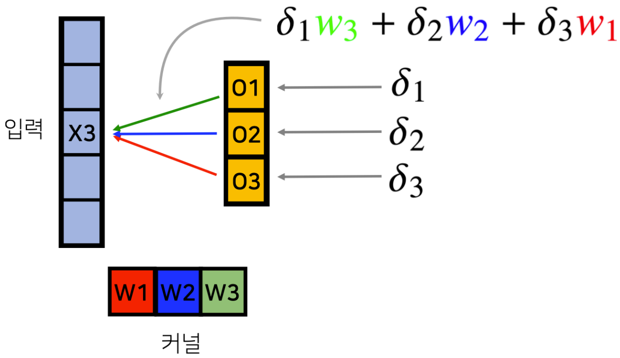
  $[(K_H, K_W, C) \ldots (K_H, K_W, C)] * (H, W, C) \to (O_H, O_W, O_C)$
  - 커널을 O_C개 사용하면 출력도 텐서가 됨
- Convolution 연산은 커널이 모든 입력데이터에 공통으로 적용됨 => 역전파를 계산할 때도 convolution 연산 나옴

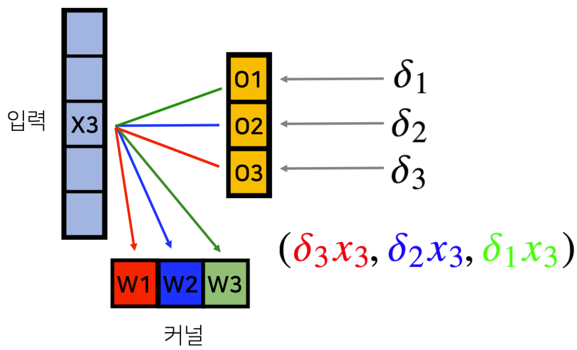
$$\begin{aligned} \frac{\partial}{\partial x}[f * g](x) &= \frac{\partial}{\partial x} \int_{\mathbb{R}^d} f(y) g(x - y) \mathrm{d}y \\ &= \int_{\mathbb{R}^d} f(y) \frac{\partial g}{\partial x}(x - y) \mathrm{d}y \\ &= [f * g'](x) \end{aligned}$$

  Discrete 일 때도 마찬가지로 성립

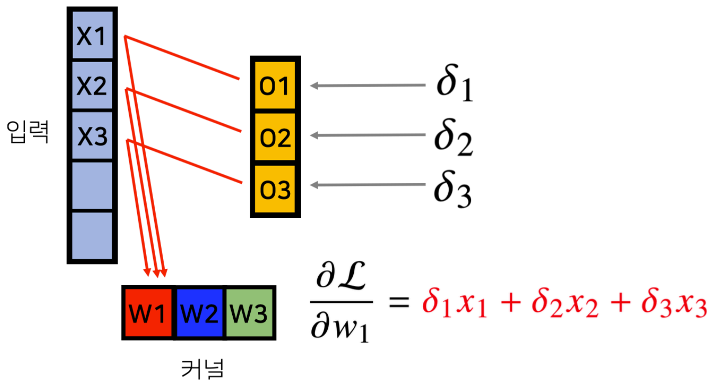


  $o_i = \sum_j w_j x_{i+j-1}$
  - 각 δ는 미분값 의미





- 역전파 단계에서 다시 커널을 통해 그레디언트 전달
- 커널에 δ에 입력값 x_3을 곱해서 전달



$\frac{\partial \mathcal{L}}{\partial w_i} = \sum_j \delta_j x_{i+j-1}$

- 각 커널에 들어오는 모든 그레디언트를 더하면 결국 convolution 연산과 같음