The Dual Form of Neural Networks Revisited

Connecting Test Time Predictions to Training Patterns via Spotlights of Attention

Kazuki Irie*



Róbert Csordás*



Jürgen Schmidhuber



* Two first authors

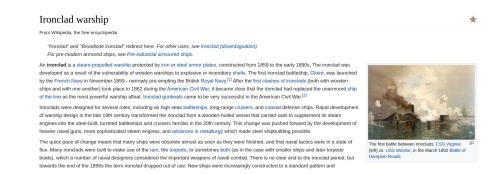




جامعة الملك عبدالله للعلوم والتقنية King Abdullah University of Science and Technology

Motivating example

- Language model trained e.g. on Wikipedia articles
- Let it do predictions on a new article



Or any other modalities you prefer... (image generation...)

Can we tell which training sentences/articles are helping its predictions?

Duality: Linear Layer vs. Attention Layer

Following two systems are equivalent

Key/Value-Attention Layer

$$oldsymbol{K} = (oldsymbol{k}_1,...,oldsymbol{k}_T) \in \mathbb{R}^{d_{ ext{in}} imes T}$$

$$oldsymbol{V} = (oldsymbol{v}_1,...,oldsymbol{v}_T) \in \mathbb{R}^{d_{ ext{out}} imes T}$$

Compute:

input
$$oldsymbol{x} \in \mathbb{R}^{d_{ ext{in}}}$$

$$S_2(\boldsymbol{x}) = \operatorname{Attention}(\boldsymbol{K}, \boldsymbol{V}, \boldsymbol{x})$$

NB: Unnormalised dot attention

$$\operatorname{Attention}(oldsymbol{K}, oldsymbol{V}, oldsymbol{q}) = oldsymbol{V} oldsymbol{K}^\intercal oldsymbol{q}$$
 vs. standard attention: $oldsymbol{V} \operatorname{softmax}(oldsymbol{K}^\intercal oldsymbol{q})$

Duality: Linear Layer vs. Attention Layer

Following two systems are equivalent

Key/Value-Attention Layer

Store:

$$oldsymbol{K} = (oldsymbol{k}_1,...,oldsymbol{k}_T) \in \mathbb{R}^{d_{ ext{in}} imes T}$$

$$oldsymbol{V} = (oldsymbol{v}_1,...,oldsymbol{v}_T) \in \mathbb{R}^{d_{ ext{out}} imes T}$$

Compute:

input
$$oldsymbol{x} \in \mathbb{R}^{d_{ ext{in}}}$$

$$S_2(\boldsymbol{x}) = \operatorname{Attention}(\boldsymbol{K}, \boldsymbol{V}, \boldsymbol{x})$$

Linear Layer

$$oldsymbol{W} = \sum_{t=1}^T oldsymbol{v}_t \otimes oldsymbol{k}_t$$

$$S_1(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{x}$$

NB: Unnormalised dot attention

Attention
$$(m{K}, m{V}, m{q}) = m{V} m{K}^\intercal m{q}$$

vs. standard attention: $m{V} \operatorname{softmax}(m{K}^\intercal m{q})$

Application to Linear Layers Trained by GD

Forward computation:

$$y = Wx$$

Backward computation (gradient descent) to update W: outer product

$$egin{aligned} oldsymbol{W}_{t+1} &= oldsymbol{W}_t igg| - \eta_t (
abla_{oldsymbol{y}} \mathcal{L})_t igg| \otimes oldsymbol{x}_t \ e_t \end{aligned}$$

for some error function $\,\mathcal{L}\,$ learning rate $\,\eta_t\,$ at step $\,t\in\mathbb{N}\,$

Application to Linear Layers Trained by GD

Forward computation:

$$y = Wx$$

Backward computation (gradient descent) to update W:

outer product

$$egin{aligned} oldsymbol{W}_{t+1} &= oldsymbol{W}_t igg[-\eta_t (
abla_{oldsymbol{y}} \mathcal{L})_t igg] \otimes oldsymbol{x}_t \ e_t \end{aligned}$$

for some error function $\,\mathcal{L}\,$ learning rate $\,\eta_t\,$ at step $\,t\in\mathbb{N}\,$

Trained weight matrix
$$oldsymbol{W} = oldsymbol{W}_0 + \sum_{t=1}^I oldsymbol{e}_t \otimes oldsymbol{x}_t$$

We can directly apply the duality from the previous slide!

Duality (Corollary)

Linear layer trained by gradient descent

Store:
$$oldsymbol{W} = oldsymbol{W}_0 + \sum_{t=1}^{T} oldsymbol{e}_t \otimes oldsymbol{x}_t$$

Compute:

$$S_1(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{x}$$

Key/Value/Attention memory storing entire training experience

$$oldsymbol{E} = (oldsymbol{e}_1,...,oldsymbol{e}_T) \in \mathbb{R}^{d_{ ext{out}} imes T}$$

$$oldsymbol{X} = (oldsymbol{x}_1,...,oldsymbol{x}_T) \in \mathbb{R}^{d_{\mathsf{in}} imes T}$$

$$S_2(\boldsymbol{x}) = \boldsymbol{W}_0 \boldsymbol{x} + \operatorname{Attention}(\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{x})$$

Matrix-version of Perceptron/Kernel Machine Duality (Aizerman et al. 1964)

Consequences/Remarks

Conceptual Remarks

- Nothing is forgotten (but it is still equivalent to a fixed storage linear layer)
- Good retrieval mechanism (capacity to selectively remember) is crucial for exploiting the memory (cf. softmax attention in Transformer)
- Etc...

Consequences/Remarks

Conceptual Remarks

- Nothing is forgotten (but it is still equivalent to a fixed storage linear layer)
- Good retrieval mechanism (capacity to selectively remember) is crucial for exploiting the memory (cf. softmax attention in Transformer)
- Etc...

Practical Consequences

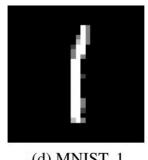
 With the dual form, we can visualise the attention from test examples (query) to training patterns (keys).

Continual Learning Experiments

- Phase 1: Train on MNIST
 - → Test accuracy of 97% on MNIST
- Phrase 2: Train on Fashion-MNIST
 - → Test accuracy of 85% on Fashion-MNIST, 45% on MNIST

But nothing is forgotten, right?

What happens exactly?



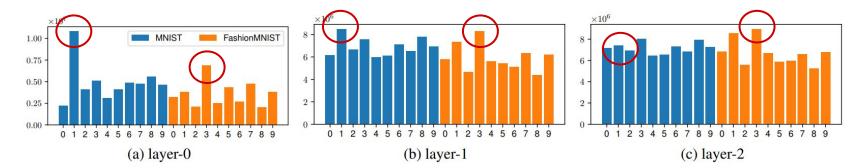




(b) F-MNIST, 3 (dress)

MNIST then F-MNIST: Attention Interference...

Example MNIST class-1 image misclassified as class 3

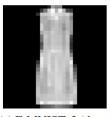


x-axis: classes (separately for MNIST and F-MNIST)

y-axis: sum of attention scores per class



MNIST class 1







(b) F-MNIST, 3 (dress)

F-MNIST class 3

Ironclad warship

(warship)



From Wikipedia, the free encyclopedia

"Ironclad" and "Broadside ironclad" redirect here. For other uses, see Ironclad (disambiguation).

For pre-modern armored ships, see Pre-industrial armoured ships.

An ironclad is a steam-propelled warship protected by iron or steel armor plates, constructed from 1859 to the early 1890s. The ironclad was developed as a result of the vulnerability of wooden warships to explosive or incendiary shells. The first ironclad battleship, Gloire, was launched by the French Navy in November 1859 - narrowly pre-empting the British Royal Navy.[1] After the first clashes of ironclads (both with wooden ships and with one another) took place in 1862 during the American Civil War, it became clear that the ironclad had replaced the unarmored ship of the line as the most powerful warship afloat. Ironclad gunboats came to be very successful in the American Civil War.[2]

Ironclads were designed for several roles, including as high seas battleships, long-range cruisers, and coastal defense ships. Rapid development of warship design in the late 19th century transformed the ironclad from a wooden-hulled vessel that carried sails to supplement its steam engines into the steel-built, turreted battleships and cruisers familiar in the 20th century. This change was pushed forward by the development of heavier naval guns, more sophisticated steam engines, and advances in metallurgy which made steel shipbuilding possible.

The quick pace of change meant that many ships were obsolete almost as soon as they were finished, and that naval tactics were in a state of flux. Many ironclads were built to make use of the ram, the torpedo, or sometimes both (as in the case with smaller ships and later torpedo boats), which a number of naval designers considered the important weapons of naval combat. There is no clear end to the ironclad period, but towards the end of the 1890s the term ironclad dropped out of use. New ships were increasingly constructed to a standard pattern and



(left) vs. USS Monitor, in the March 1862 Battle of

Ironclad warship (warship)

From Wikipedia, the free encyclopedia

"Ironclad" and "Broadside ironclad" redirect here. For other uses, see Ironclad (disambiguation).

For pre-modern armored ships, see Pre-industrial armoured ships.

An ironclad is a steam-propelled warship protected by iron or steel armor plates, constructed from 1859 to the early 1890s. The ironclad was developed as a result of the vulnerability of wooden warships to explosive or incendiary shells. The first ironclad battleship, Gloire, was launched by the French Navy in November 1859 - narrowly pre-empting the British Royal Navy.¹¹ After the first classhes of ironclads (both with wooden ships and with one another) took place in 1862 during the American Civil War, it became clear that the ironclad had replaced the unarmored ship of the line as the most powerful warship afficial. Ironclad gunboats came to be very successful in the American Civil War.^[2]

Ironclads were designed for several roles, including as high seas battleships, long-range cruisers, and coastal defense ships. Rapid development of warship design in the late 19th century transformed the ironclad from a wooden-hulled vessel that carried sails to supplement its steam engines into the steel-built, turreted battleships and cruisers familiar in the 20th century. This change was pushed forward by the development of heavier naval guns, more sophisticated steam engines, and advances in metalfurgy which made steel shipbuilding possible.

The quick pace of change meant that many ships were obsolete almost as soon as they were finished, and that naval tactics were in a state of flux. Many inonclads were built to make use of the ram, the torpedo, or sometimes both (as in the case with smaller ships and later torpedo boats), which a number of naval designers considered the important weapons of naval combat. There is no clear end to the ironclad preiod, but towards the end of the 1990s the term ironclad dropped out of use. New ships were increasingly constructed to a standard pattern.

Test.



The first battle between ironclads: CSS Virginia (left) vs. USS Monitor, in the March 1862 Battle of Hampton Roads

Portuguese ironclad Vasco da Gama

From Wikipedia, the free encyclopedia.

Vasco d. Cames was an inoxidation of the Portuguese Naivy Usal in the 1970s by the Thames ton Works in Lindonic Ordered is strengthen the deletered of the Portuguese capital of Libbon, Veroco de Cames was large-handed in 1976 to de completed in 1978. This enterior is 1978. This enterior is 1978 to 1979 to 1979 to 1979, West of the Puruguese Boot for the majority of the long and praceful career. She was rebuilt and heavy modernized between 1901 and 1903. Her crow was reviewed in evoids in 1913 and 1914, during the later event, they benthereful Lisbon and killed around one hundred people. Long-aince obsolete by the 1905s, Vestor do Games was fairly and for scapping in 1934.

Contents [hide]
1 Design
1.1 Modifications
2 Service history
3 Notes
4 References

(warship)

Design [edit]

Vasco de Gama was the only capital ship to be built for the Portuguese Navy; ordered from a British shipyard, she was intended to defend the applied of Eliston from second attracts [1]



Canal in 1895
History
Portugal

Train

SMS Markgraf

Errom Wikinertia, the free encyclonertia

SMS Managard² is our bre held trained in the media picking class. She served in the imposful German Navy Arrang World Wir I. The battankly used laid down in Neverther 1951 and launched on 4 Jave 1913. She was formally commissioned into the imposful Navy on 1. October 1914, just over two motifs and the two outbreak of war in Europe. Markgard was arread with two 35-certification (12.00 inj gars in New Initial matries and could stem at a top speed of 21 locot, (26 losin); 28 injoy). Markgard was named in hours of the royal family of Boden. The name Markgard is a rank of German notibilly and a capacided to the England Happarce of Manages.

Along with the three states rating, Kingle, Clossor Kurlists, and Occopionis, Manigar favis layer in most of the fined anticon during the war, including the Bastic of Juliand on 31 May and 1 Ame 2016. All Juliand, Marling of uses the ritid ship in the German in and heavily engaged by the opposition, the film Closed Faster in the sustained film large-periality that was the cross workered 22 cassalities. Marling all along participated in Operation Allons, the compared of the Guilf of Eggs, in lise 1917. The stip was damaged by a mine white on route to Germany following the successful conclusion of the opposition.

After Germany's defect in the war and the signing of the American in November 1918. Modify and on our of the capital indips of the high Seas Feet ware interned by the Popul Nay in Scass, Disc. The sibes ware destined and reduced to sidned concern while the Aller allows regulated the final version of the Treaty of Versallers. On 21. June 1918, days before the treaty was signed, the commander of the interned feet. Rear Administ Louising van Richart, ordered at the feet to be sociated to execute that the Bellin Andre feet also is side the along. Unlike most of the sociated object, Marging was never maked to except perior the version stall staffing on the bettern of the version of the sociated object, Marging was never maked to except perior the version stallaring on the bettern of the sociated object. Marging was never maked to except perior the version stall staffing on the bettern of the version of the sociated object. Marging was never maked to except perior the version of the sociated object. Marging was never maked to except perior the version of the sociated object. Marging was never maked to except the version of the sociated object. Marging was never maked to except the version of the sociated object. Marging was never maked to except the version of the sociated object. Marging was never maked to except the version of the sociated object. Marging was never maked to except the version of t



Recognition develop of a KVing-class betreen History

History

German Empire

Builder AG Wester, Bremen
Laid down Neverriber 1931
Lassrched 4 June 1933

Commissioned 1 October 1954
Falte Southee 21, June 1931 in Guster

(warship)

Italian cruiser Aretusa

From Wikipedia, the free encyclopedia.

For other ships with the same name, see Italian ship Aretusa.

Aretura was a torpedo custer of the Parterope class built for the fallatin Regul Marriar (Royal Boson) in the 1800s. Laid door in 3.we 1809 at the Cantern revisit in India Chiefact Brighty Artification (Partern Royal Boson) in the 1800s. Laid door in 3.we 1809 at the Royal Royal Boson (Partern Royal Boson Royal Royal Boson Royal Boson Royal Royal Boson Royal Royal

Contents [hide]
1 Design
2 Service history
2.1 Italo-Turkish War
3 Footnotes
3.1 Notes
3.2 Citations
4 References
5 External links

Train ®



(warship)

(warship) Ironclad warship

From Wikipedia, the free encyclopedia

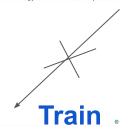
"Ironclad" and "Broadside ironclad" redirect here. For other uses, see Ironclad (disambiguation).

For pre-modern armored ships, see Pre-industrial armoured ships.

An ironclad is a steam-propelled warship protected by iron or steel armor plates, constructed from 1859 to the early 1890s. The ironclad was developed as a result of the vulnerability of wooden warships to explosive or incendiary shells. The first ironclad battleship. Gloire, was launched by the French Navy in November 1859 - narrowly pre-empting the British Royal Navy.[1] After the first clashes of ironclads (both with wooden ships and with one another) took place in 1862 during the American Civil War, it became clear that the ironclad had replaced the unarmored ship of the line as the most powerful warship affoat, Ironclad gunboats came to be very successful in the American Civil War.[2]

Ironclads were designed for several roles, including as high seas battleships, long-range cruisers, and coastal defense ships. Rapid development of warship design in the late 19th century transformed the ironclad from a wooden-hulled vessel that carried sails to supplement its steam engines into the steel-built, turreted battleships and cruisers familiar in the 20th century. This change was pushed forward by the development of heavier naval guns, more sophisticated steam engines, and advances in metallurgy which made steel shipbuilding possible.

The guick pace of change meant that many ships were obsolete almost as soon as they were finished, and that naval tactics were in a state of flux. Many ironclads were built to make use of the ram, the torpedo, or sometimes both (as in the case with smaller ships and later torpedo boats), which a number of naval designers considered the important weapons of naval combat. There is no clear end to the ironclad period, but towards the end of the 1890s the term ironclad dropped out of use. New ships were increasingly constructed to a standard pattern and



Nina Simone

Eunice Kathleen Waymon (February 21, 1933 – April 21, 2003), known professionally as Nina Simone, was an American singer, songwiter, plants, and civil rights activist. Her music spanned styles including classical, folk, pospet, blues, jazz, R&B, and pop-

The sixth of eight children born to a poor family in Tryon, North Carolina, Simone initially aspired to be a concert planist [1] With the help of a few Institute of Music in Philadelphia, where she was denied admission despite a well received audition [2] which she attributed to racium, in 2002, and day

To make a living, Simone started playing plans at a nightdub in Atlantic City. She changed her name to "Nina Simone" to disguise herself from family members, having chosen to nity "the devil's music" or so called "cocktall plans". She was told in the night-high that she would have to sing to her own making her debut with Little Girl Rive. She hard a hit single in the United States in 1958 with T.L. russ Voy. Premit Till Her musical state fused ancord and pop with classical music. In particular Johann Sebastian Bach, ⁽⁶⁾ and accompanied expressive, jazz-like singing in her controllo voice, ⁽⁷⁸⁾

(singer)



1.3 1959-1954: Burgeoning popularit

1.5 1974-1993 Later Re-1.6 1993-2003: Final years, illness and death

Test.



The first battle between ironclads: CSS Virginia (left) vs. USS Monitor, in the March 1862 Battle of Hampton Roads

Portuguese ironclad Vasco da Gama

From Wikipedia, the free encyclopedia

Vision de Game was an invariant of the Portuguese New helt in the 1870s by the Thomas Iron Works in London Ordered to strengthen the defenses of the Portuguese capital of Lisbon, Vasco da Gama was launched in 1876 and completed in 1878. She served as the flagship of the Portuguese fleet for the majority of her long and peaceful career. She was rebuilt and heavily modernized between 1901 and 1903. Her crew was involved in results in 1913 and 1914; during the latter event, they hombarded Lisbon and killed around one hundred people. Long since obsolete by the 1930s, Vasco da Gama was finally sold for scrapping in 1935.

1 Design 1.1 Modification 2 Service history

(warship)

Design [edit]

Vasco de Gama was the only capital ship to be built for the Portuguese Navy; ordered from a British shipvard, she was intended to defend the



Train

SMS Markgraf

Errom Wikingrija, the free encycloneria

SMS Markora(fill was the third battleship of the four-ship Könip class. She served in the Imperial German Nevy during World War I. The battleship was laid down in November 1911 and launched on 4 June 1913. She was formally commissioned into the Imperial Navy on 1 October 1914, just over two months after the outbreak of war in Europe. Markgraf was armed with ten 30.5-centimeter (12.0 in) guns in five twin turrets and could steam at a top speed of 21 knots (39 km/h; 24 mph). Markoral was named in honor of the royal family of Baden. The name Markoral is a rank of German nobility and is equivalent to the English Margrave, or Marquess

Along with her three sister ships, König, Grosser Kurfürst, and Kronprinz, Markgraf took part in most of the fleet actions during the war, including the Battle of Jutland on 31 May and 1 June 1916. At Jutland. Markgraf was the third ship in the German line and heavily encaged by the opposin British Grand Elect she sustained five large-caliber hits and her resw suffered 23 casualties. Markins also narticipated in Operation Althory the conquest of the Gulf of Rica, in late 1917. The ship was damaged by a mine while en route to Germany following the successful conclusion of the operation

After Germany's defeat in the war and the signing of the Armistice in November 1918, Markaraf and most of the capital ships of the High Seas Fixet were interned by the Royal Navy in Scapa Flow. The ships were disarmed and reduced to skeleton crews while the Allied powers negotiated the final version of the Treaty of Versailles. On 21 June 1919, days before the treaty was signed, the commander of the interned fleet. Rear Admiral Ludwig von Reuter, ordered the fleet to be scuttled to ensure that the British would not be able to seize the ships. Unlike most of the scuttled ships. Markoraf was never raised for scrapping; the wreck is still sitting on the bottom of the bay.



November 1911 Laid down Launched 4 June 1913 oned 1 October 1914 Scuttled 21 June 1919 in Gutter

(warship)

Italian cruiser Aretusa

From Wikipedia, the free encyclopedia

For other ships with the same name, see Italian ship Aretusa.

Aretusa was a torpedo cruiser of the Partenope class built for the Italian Regia Marina (Royal Navy) in the 1880s. Laid down in June 1889 at the Cantiere payale fratelli Orlando shipyard, she was launched in March 1891 and was commissioned in September 1892. Her main armament were her six torpedo tubes, which were supported by a battery of ten small-caliber guns. Aretusa spent most of her career in the main Italian fleet. where she was primarily occupied with training exercises. At the start of the Italo-Turkish War in September 1911, she was assigned to the Red Sea Squadron in Italian Eritrea. She bombarded Ottoman positions in the Arabian Peninsula and took part in a blockade of the coast. Worn out by the end of the war in October 1912, Aretusa was sold for scrap that December and broken up.



Train



(warship)



Contents (line)

1. Design
1. Stores Abstract heavy
2. Roderwards
2. Service heavy
3. Roder
4. Roderwards
4. Roderwards

Vesco de Gansa was the cuty capital ship to be built for the Portuguese Nany, ordered from a British shippard, she was intended to defend the

SMS Markgraf

From Villopeta, the has encyclopeda

3. SMS Markgraf

From Villopeta, the has encyclopeda

3.

Portuguese ironclad Vasco da Gama

by the 1930s, Vasco da Gama was finally sold for scrapping in 1935.

Vasco dis Cama was an inonistid of the Portiquiese Nally bull in the 1870's by the Thaines Iron Vivols in London. Ordered to strengthen the defenses of the Portiquiese capital of Little, Visco dis Camar was submished in 1879 and completed in 1879. She served as the fleaght for the Portiquiese feet for the majority of the Iron and paradic camer. She was rebuilt and heavily modernized between 1901, and 1903, Her crew waster in the contract of the Iron and Iron

Query (Ironclad warship)

... Her **principal** role was for combat in the English Channel and other European ...

Top-1 (*Portuguese ironclad*) Top-2 (*SMS Markgraf*)

Top-3 (Italian cruiser Aretusa)

Her sailing rig also was removed. Her **main** battery guns were replaced with new between the two funnels. Her **secondary** armament consisted of fourteen 15 cm ... single mounts. Her **primary** offensive weapon was her five 450 mm.

Negative (*Nina Simone*)

... written especially for the singer. Her **first** hit song in America was her rendition ...



Lamenter novine trainers Liveanous sorgiums, see was sustinction in nation, sex and was commissionate in separative Intil. Internal inflamment were here is to repose to hisse, which were supported by a battery of the made callenge grans. Anether specer to the main latitude feet, where she was primarily occupied with training exercises. At the start of the tato-Turkish War in September 1911, she was assigned to the Red Sea Seguadron in Intelna Entress. She bombarded Ottoman positions in the Arabban Periminal and took part in a blockade of the coast. Wom out by the end of the war to Cookber 1912, Avertosa was sold for scorp that December and troken up.

Contentes pixel.

Design
2 Sentice history
2. Illied-Turkish War
3 Footones
3.1 Notice
3.2 Collabors
4 References
5.5 Estemail lats



(warship)

Thank you for your attention.

https://github.com/robertcsordas/linear_layer_as_attention

Hope to see you at the poster!